Fach – Sprache – Kommunikation 2

Andrea Hunziker Heeb

Effects of directionality on professional translators' processes, products and effort

UV Universitätsverlag Hildesheim

Andrea Hunziker Heeb

Effects of directionality on professional translators' processes, products and effort

Fach – Sprache – Kommunikation

herausgegeben von
Franziska Heidrich-Wilhelms und Klaus Schubert

Band 2

Andrea Hunziker Heeb

Effects of directionality on professional translators' processes, products and effort

Universitätsverlag Hildesheim

2020

Andrea Hunziker Heeb

Effects of directionality on professional translators' processes, products and effort

Universitätsverlag Hildesheim

2020

Das Dokument steht im Internet kostenfrei als elektronische Publikation (Open Access) zur Verfügung unter: https://dx.doi.org/10.18442/136

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Satz, Layout und Titelblattgestaltung: Jan Jäger
Titelbild: ThreeMilesPerHour auf Pixabay

www.uni-hildesheim.de/bibliothek/universitaetsverlag/
ISSN 2699-7118

Vom Fachbereich 3 (Sprach- und Informationswissenschaften) der Universität Hildesheim zur Erlangung des Grades einer Doktorin der Philosophie (Dr. phil.) angenommene Dissertation von

Andrea Sabine Hunziker Heeb

Gutachter: Prof. Dr. Klaus Schubert (Universität Hildesheim)
Gutachterin: Prof. Dr. Nathalie Mälzer (Universität Hildesheim)

Tag der mündlichen Prüfung: 15. August 2019

Hil2

Zusammenfassung

Diese empirische Arbeit in der kognitiven Translationswissenschaft geht der Frage nach, ob und wie sich das Übersetzen in eine Fremdsprache (L2) hinsichtlich des Arbeitsprozesses, des Aufwands und des Ergebnisses vom Übersetzen in die Muttersprache (L1) unterscheidet. Anlass dazu gab der Umstand, dass gegenüber dem Übersetzen in die L2 in der Übersetzungswissenschaft, Lehre und Berufswelt immer noch Vorurteile bestehen, obwohl diese Übersetzungsrichtung eine weit verbreitete und für manche Sprachgemeinschaften sogar unverzichtbare Praxis darstellt. Die Studie konzentriert sich auf das Sprachenpaar Deutsch-Englisch, auch im Hinblick auf die anhaltend hohe Nachfrage nach Übersetzungen ins Englische im deutschsprachigen Raum.

Unter kontrollierten Bedingungen führten FachübersetzerInnen, die beruflich sowohl aus ihrer L1 Deutsch in ihre L2 Englisch als auch umgekehrt arbeiten, Übersetzungsaufträge aus. Als Kontrollgruppe fungierten FachübersetzerInnen, die ausschliesslich in ihre L1 übersetzen, d. h. entweder ins Deutsche oder ins Englische.

Gestützt auf den Mixed-Methods-Ansatz wurden in der Hauptuntersuchung die Übersetzungsprozesse mit Methoden der Translationsprozessforschung – d. h. mit Protokollierung der Tastenanschläge, Aufnahme des Computerbildschirms, Protokollierung retrospektiven Lauten Denkens, Blickerfassung und Interview – erfasst und ausgewertet. Abgeleitet von den Haupttätigkeiten Schreiben, Überarbeiten, Recherchieren und Pausieren wurden zahlreiche Prozessdeskriptoren analysiert und verglichen. Durch Triangulation mit Daten aus der Zieltextanalyse wurden Indikatoren für den Übersetzungsaufwand gemessen. Die Auswertung ergab für die überwiegende Mehrheit der Prozessdeskriptoren und der Aufwandindikatoren keine statistisch signifikanten Unterschiede aufgrund der Übersetzungsrichtung.

In der Nebenuntersuchung beurteilten intendierte Adressatinnen und Adressaten der englischen Zieltexte deren Angemessenheit anhand eigener Kriterien und ohne Rückgriff auf den Ausgangstext. Die Auswertung ergab keinen signifikanten Unterschied in der Zieltextqualität aufgrund der Übersetzungsrichtung.

Somit konnte diese Studie zeigen, dass die Übersetzungsrichtung offensichtlich keine bedeutsame Rolle beim Übersetzungsprozess, der Produktqualität und dem übersetzerischen Aufwand spielt. Damit entkräftet sie vorgebrachte Vorbehalte gegen das Übersetzen in die L2.

Abstract

This empirical study in Cognitive Translation Studies investigates whether and how translation into a foreign language (L2) differs from translation into the native language (L1) in terms of the translation process, the effort and the product. The rationale behind the study was to investigate prejudices towards translation into the L2 that persist in Translation Studies, translation training and practice despite the fact that it is widely practiced and even is the main translation direction in many language communities. The study focuses on the German-English language pair, also in view of the continually high demand for translations into English in German-speaking countries.

As participants, qualified non-literary translators who are used to translating professionally both from their L1 German into their L2 English and vice versa performed two translation tasks under controlled conditions. The control groups consisted of professional translators who translate exclusively into their L1, i. e. either into German or into English.

Following a mixed methods approach, the main study investigated the translation processes using methods of translation process research: keystroke logging, screen recording, retrospective verbal protocol, eye tracking and interview. From the principal process activities writing, revising, information searching and pausing, process descriptors were derived, analysed and compared. By triangulating results with data from the target text analysis, indicators of translation effort were measured. For the vast majority of the process descriptors and effort indicators, the results showed no statistically significant differences with regard to translation direction. In the substudy, intended addressees of the English target texts assessed the texts' acceptability based on their own criteria and without resorting to the source text. The results showed no statistically significant difference in target text quality with regard to translation direction. In summary, this study showed that translation direction clearly plays no significant role in the translation process, the product quality and the translation effort. It thus invalidates preconceptions about translation into L2.

Acknowledgements

I would like to thank everyone who accompanied and supported me on this journey, whether on a small section or on the whole route. I would also like to thank those people who only crossed my path and still took an interest in my work.

To my family, friends, study participants, supervisors, past and present office colleagues, the IUED Translation Studies team, superiors and colleagues at the ZHAW School of Applied Linguistics, fellow PhD students in the Doktorandennetzwerk at the Universität Hildesheim, fellow researchers of the TREC network and the Swiss National Science Foundation.

Special thanks go to Maureen and Patrick: You've rocked my world! Should I ever forget what makes the world go round, please remind me: It's translation, stupid!

Winterthur, March 2019

Table of contents

1 Introduction 19

2 Research questions and scope 25

3 Theoretical underpinnings and literature review 29
3.1 The concept of directionality in translation 32
3.2 Directionality, translation processes and products in Cognitive Translation Studies 38
3.3 Evaluation of translation product quality for directionality studies 44
3.4 The concept of translator effort 48
3.5 Directionality and translator effort in Cognitive Translation Studies 59
3.6 Directionality in Interpreting Studies 64
3.7 Conclusions from the review of the literature 65

4 Methodology 69
4.1 Overall study design: Mixed methods approach 69
4.2 Main study: Directionality and the translation process 71
4.2.1 Main study design 71
4.2.2 Main study participants: Professional translators 73
4.2.3 Main study stimuli and tasks: Source texts and translation briefs 76
4.2.4 Main study data collection methods 77
4.2.5 Main study data collection procedure 80
4.2.6 Main study operationalisations 81
4.2.7 Main study ethical issues: Professional translators as participants 85
4.2.8 Main study methodological issues 86
4.3 Substudy: Directionality and the translation product 89
4.3.1 Substudy design 89
4.3.2 Substudy participants: Text users as evaluators 90
4.3.3 Substudy stimuli and task: English target texts and their evaluation 92
4.3.4 Substudy data collection methods 93
4.3.5 Substudy data collection procedure 94
4.3.6 Substudy operationalisations 95
4.3.7 Substudy ethical issues: Students as participants 97
4.3.8 Substudy methodological issues 98

4.4 Directionality and translator effort 98
4.4.1 Character production effort 100
4.4.2 Revision effort 100
4.4.3 Information search effort 100
4.4.4 Coordination effort 100
4.4.5 Mean fixation duration 101
4.5 Relations between characteristics of professional translation..... 101

5 The translators' processes 103
5.1 Global process measures 104
5.1.1 Process duration 104
5.1.2 Character count 106
5.1.3 Summary of results for global process measures 107
5.2 General description of the translation processes 108
5.2.1 Total number of process activities 109
5.2.2 Proportion of activity types in the translation processes ... 110
5.2.3 Summary of results for the general translation processes measures..... 114
5.3 Writing in the translation process 114
5.3.1 Total number of writing incidents 114
5.3.2 Frequency of writing 115
5.3.3 Mean duration of writing incidents 117
5.3.4 Mean length of writing incidents 118
5.3.5 Summary of results for the writing measures 119
5.4 Revising in the translation process 119
5.4.1 Total number of revisions 120
5.4.2 Frequency of revising 120
5.4.3 Proportion of revisions per translation process phase 122
5.4.4 Mean length of revisions in revision phase 124
5.4.5 Summary of results for the revising measures 126
5.5 Information seeking in the translation process 126
5.5.1 Total number of information searches 127
5.5.2 Frequency of information seeking 128
5.5.3 Mean duration of information searches 129
5.5.4 Ratio of single-step to multi-step information searches 130
5.5.5 Types of online resources used 132
5.5.6 Summary of results for the information seeking measures. 135
5.6 Pausing in the translation process 135
5.6.1 Total number of pauses 136
5.6.2 Frequency of pausing 136
5.6.3 Mean duration of pauses 138
5.6.4 Summary of results for the pausing measures 139

5.7 Summary of results from the translators' processes 139
5.7.1 Intragroup comparisons 139
5.7.2 Intergroup comparisons 140

6 The translators' products 143
6.1 Global measures of the translation products 143
6.1.1 Number of target text characters 144
6.1.2 Number of target text words 144
6.1.3 Productivity 145
6.2 Quality of the translators' products 146
6.2.1 Readability of the English target texts 147
6.2.2 Acceptability of the English target texts 149
6.2.3 Raters' evaluation criteria 151
6.2.4 Correlation of acceptability and readability 152
6.2.5 Sequence of the translation tasks and product quality 153
6.3 Summary of results from the translators' products 153

7 The translators' effort 155
7.1 Character production effort 156
7.2 Revision effort 158
7.3 Information search effort 159
7.4 Coordination effort 161
7.5 Mean fixation duration 163
7.6 Summary of results from the translators' effort 164

8 Relations between characteristics of professional translation 167

9 Synthesis and outlook 173
9.1 Answers to my research questions 173
9.2 Discussion of the study results 183
9.3 Limitations of the study 187
9.4 Relevance of the study and proposition for reconceptualisation 189
9.5 Suggestions for further research 192
9.6 Potential implications of the study's findings 193

10 Bibliography 197

11 Appendices 219

List of figures

Figure 1: Overall study design: mixed methods approach 70
Figure 2: Group comparisons in main study 73
Figure 3: Usability lab setting with participant in the background and observers in the foreground 79
Figure 4: Instructions for evaluators of English TTs 93
Figure 5: Template of ranking list for raters 94
Figure 6: Examples of a bidirectional translator's processes into L2 and L1 (Bidir2_L2 and Bidir2_L1) as sequences of the four main process activities 108
Figure 7: Proportions for no. of activity types in Bidir_L2, Bidir_L1, UniEnglish_L1 and UniGerman_L1 processes (%) 111
Figure 8: Proportions for duration of activity types in Bidir_L2, Bidir_L1, UniEnglish_L1 and UniGerman_L1 processes (%) 113
Figure 9: Frequency of writing in Bidir_L2 and Bidir_L1 processes ordered by translator (no./10 min) 116
Figure 10: Mean duration of writing incidents (sec) for Bidir_L2 and Bidir_L1 processes, ordered by translator 117
Figure 11: Mean length of writing incidents (char) for Bidir_L2 and Bidir_L1 processes, ordered by translator 118
Figure 12: Frequency of revising in Bidir_L2 and Bidir_L1 processes ordered by translator (no/10 min) 121
Figure 13: Frequency of revising in UniEnglish_L1 and Bidir_L2 processes in ascending order (no/10 min) 122
Figure 14: Proportion of revisions per TP phase (%) in Bidir_L2 and Bidir_L1 processes, ordered by translator 123
Figure 15: Mean length of deletions and insertions (in characters) in the revision phase of Bidir_L2 and Bidir_L1 processes, ordered by translator 125
Figure 16: Frequency of information seeking in Bidir_L2 and Bidir_L1 processes (no/10 min), ordered by translator 128
Figure 17: Mean duration of information searches (sec) for Bidir_L2 and Bidir_L1 processes, ordered by translator 129
Figure 18: Ratio of single-step to multi-step information searches (%) in Bidir_L2 and Bidir_L1 processes ordered by translator 131
Figure 19: Ratio of single-step to multi-step information searches (%) in UniGerman_L1 and Bidir_L1 processes, in ascending order by single-step searches 132

Figure 20: Mean use of types of online resources in Bidir_L2 and Bidir_L1 processes (%) 133
Figure 21: Mean use of types of online resources in Bidir_L2 and UniEnglish_L1 processes (%) 134
Figure 22: Mean use of types of online resources in Bidir_L1 and UniGerman_L1 processes (%) 134
Figure 23: Frequency of pausing in Bidir_L2 and Bidir_L1 processes ordered by translator (no./10 min) 137
Figure 24: Frequency of pausing in UniEnglish_L1 and Bidir_L2 processes in ascending order (no./10 min) 137
Figure 25: Mean duration of pauses (sec) for Bidir_L2 and Bidir_L1 processes, ordered by translator 138
Figure 26: Productivity for the bidirectional translators for the translation processes into L2 and into L1, ordered by translator (words/min) 145
Figure 27: Productivity for UniEnglish_L1 and Bidir_L2 groups, in ascending order (words/min) 146
Figure 28: Productivity for UniGerman_L1 and Bidir_L1 groups, in ascending order (words/min) 146
Figure 29: Flesch Reading Ease scores for English TTs of Bidir_L2 and UniEnglish_L1 groups, in ascending order 147
Figure 30: Acceptability scores for TTs of Bidir_L2 and UniEnglish_L1 in ascending order, based on evaluation by total rater group (n=36) 149
Figure 31: Acceptability scores for TTs of Bidir_L2 and UniEnglish_L1 based on evaluation by L1 raters and L2 raters, (in ascending order, based on evaluation by total rater group) . 150
Figure 32: Character production effort for Bidir_L2 and Bidir_L1, ordered by translator (process char/TT char) 157
Figure 33: Revision effort for Bidir_L2 and Bidir_L1, ordered by translator (revisions/10 TT words) 158
Figure 34: Information search effort for Bidir_L2 and Bidir_L1, ordered by translator (queries/10 TT words) 160
Figure 35: Information search effort for UniEnglish_L1 and Bidir_L2, in ascending order (queries/10 TT words) 160
Figure 36: Coordination effort for Bidir_L2 and Bidir_L1 processes (%), ordered by translator 162
Figure 37: *Mean fixation duration* (ms) for Bidir_L2 and Bidir_L1 processes, ordered by translator 163

List of tables

Table 1: Quasi-experimental design of main study: Participant groups and translation tasks 72
Table 2: Overview of data collected in main study 77
Table 3: Rater groups in target text evaluation task 91
Table 4: Overview of data collected in substudy 93
Table 5: Scores assigned to the positions in the ranking task to determine product acceptability 96
Table 6: Indicators of translator effort and their components 99
Table 7: Duration of Bidir_L2 and Bidir_L1 translation processes (in seconds) ordered by translator 104
Table 8: Duration of UniEnglish_L1 and Bidir_L2 processes in ascending order (in seconds) 105
Table 9: Duration of UniGerman_L1 and Bidir_L1 processes in ascending order (in seconds) 106
Table 10: Character count for Bidir_L2 and Bidir_L1 processes (no.) ordered by translator 106
Table 11: List of all the measures and the results for intergroup comparisons, in sequential order as reported in this chapter 141
Table 12: Self-reported evaluation criteria per rater group (%), in descending order by total rater group 152
Table 13: Rank correlations between readability and product quality of the English target texts (Kendall's τ) 153
Table 14: Overview of comparisons of translator effort indicators (significance level indicated; otherwise non-significant) 164
Table 15: Representation of correlation coefficient ranges 167
Table 16: Correlations between *translator effort*, *productivity* and *professional experience* for the bidirectional translators 168
Table 17: Correlations between translator effort, productivity, professional experience and product quality for the Bidir_L2 and the UniEnglish_L1 groups 169
Table 18: Correlations between translator effort, productivity and professional experience for the Bidir_L1 and the UniGerman_L1 groups 170

List of abbreviations

Bidir	bidirectional translators working into and out of German, their first language
CAT	computer-aided translation
CTS	Cognitive Translation Studies
ELF	English as a lingua franca
ET	eye tracking
KSL	keystroke logging
MT	machine translation
NMT	neural machine translation
RVP	cued retrospection
SCR	screen recording
SL	source language
ST	source text
TL	target language
TP	translation process
TPR	translation process research
TT	target text
UniEnglish	unidirectional translators working exclusively into English, their first language
UniGerman	unidirectional translators working exclusively into German, their first language

1 Introduction

Translation comes in many forms and guises. A common denominator is that it involves the rendering of speech or text in another language. While this act is probably performed by many people every day, translation is also a profession. Professional translators mostly work with non-literary written material and at least two languages. Usually, one of these working languages is their so-called native or first language. In many countries in Western Europe, translation into the first language (L1 translation) is the main translation direction, while translation into a second language (L2 translation) is also widely practised. It seems however, that L1 translation is considered the gold standard. This quality claim has found its way onto websites of translation agencies and freelancers as the so-called *mother tongue principle*[1], at least in countries such as Switzerland. This suggests that translating exclusively into one's L1 would be a criterion of professionalism and guarantee services of high quality. As if the mode of how a language is acquired could ensure its proficient use in translation (Pokorn 2005, 6). Another indication of prejudice against L2 translation is that L2 translators may only have restricted access to professional bodies in some countries. In addition, training in L2 translation at university level often seems to resemble language learning exercises rather than preparation for professional practice. Anecdotal evidence about L2 translation producing inherently low quality and involving disproportionately large effort has perpetuated the view that the direction into L1 is preferred.

In Translation Studies, reservations towards L2 translation are reflected in a scarcity of studies so far (Apfelthaler 2013). This underrepresentation also ignores the continually increasing demand for translations into English as a means of global communication. Simple arithmetic suggests that the demand cannot be met by L1 translators alone. In the context of German-speaking countries, L2 translation into English is probably the version most practised and generally accepted by the market. The object of the empirical study presented here is the real-world phenomenon of professional translation into a second language. The study aims to add to the understanding of how non-literary, i. e. specialised L2 translation from German into English is actually performed by professional L2 translators. It does this by investigating the translation processes, the translation products

1 Often, the expressions *mother tongue principle* (*Muttersprachprinzip* in German) or *native-speaker translator* (reported for the Czech Republic and Germany in Šebesta 2013) are used without further explanation, which does not make sense since every translator is indeed a speaker of a native language or mother tongue.

and the effort for translation of formally trained or accredited professional translators who are used to translating back and forth between German and English. The translators are monitored while they are performing translation tasks and their activities are captured without interfering with their usual procedures. The analysis of those activities then provides information about features of the translation process and also the translators' expended effort, especially but not only cognitive effort. As the latter can only be measured indirectly, it makes sense to collect data from different sources and determine more than one indicator. In order to be able to make solid comparisons and to determine whether differences in processes, products or effort are related to directionality, that is the translation direction, translators who are used to translating exclusively into their L1 (German or English, respectively) are also included. According to their translation direction(s), the study participants translate the same texts along identical task requirements in a controlled setting. This will be done in the main study. In the substudy, the quality of the produced texts is then assessed. The assessments are performed by users of those texts who apply their own criteria, just as in real life.

Assumptions about the wide distribution of L2 translation have been supported by several surveys. In a small-scale survey among freelance translators who operate in the Swiss market, almost thirty percent of the respondents stated that they engaged both in L1 and in L2 translation (den Dekker 2017, 24). Although those results may not be representative, they show that L2 translation is professionally practised in Switzerland. With regard to its wider distribution, an international online survey initiated by *The International Association of Professional Translators and Interpreters* shows that over 50 % of the 772 respondents also translate into their L2 (IAPTI's Ethics Committee 2015, 13). In survey by Schmitt, Gerstmeyer and Müller (2016, 171) on behalf of CIUTI[2] among graduates from its member institutions, about 20 % of the responding translators said that they mainly work into their L2.[3] In yet another international survey with 1,850 respondents from 50 countries, 43 % reported that they also translate into their L2 (Ehrensberger-Dow et al. 2016, 5). When analysed by employment position, significantly more commercial and freelance translators worked into their L2 than institutional translators did (Ehrensberger-Dow et al. 2016, 5)[4], confirming the impression that L2 translation is needed by the market.

2 CIUTI: Conférence internationale permanente d'Instituts universitaires de Traducteurs et Interprètes. CIUTI is an international association of training institutions for translators and interpreters that adhere to certain standards.

3 In total, 923 translators and interpreters answered this question.

4 48 % of the commercial translators and 45 % of the freelancers vs. 33 % of the institutional translators; $p<0.001$.

Country-specific surveys or accounts show that L2 translation is practised in, for example, Denmark (Lorenzo 2003), Slovenia (Pokorn 2008), Croatia (Pavlović 2007b) and Poland (Whyatt and Kościuczuk 2013). While these countries can be defined as communities with languages of limited diffusion that may be in greater need for L2 translation,[5] other countries with L2 translation practice, such as Spain (Kelly, Nobs, Sánchez and Way 2003), China (B. Wang 2011) and Brazil (Ferreira 2010), cannot.[6]

At a conference on L2 translation training and practice in Europe in 2015, the vice president of the German Federal Association of Interpreters and Translators (BDÜ) defended the interests of the association's more than 7,500 members. She stressed the importance of L2 translators into English for the German automotive industry and urged translation training institutions to adequately prepare their students (Groethuysen 2015). Apparently, important industry sectors in Germany rely heavily on German-speaking translators working into English who are experts in their subject fields. L2 translation into English thus seems to be in demand. Moreover, clients' focus is on translators understanding their business needs, and they trust them to be able to work in both translation directions (IAPTI's Ethics Committee 2015, 31; Whyatt and Kościuczuk 2013, 73). More than 70 % of the members of the BDÜ who offer English-German translation services into their L1 also offer German-English into their L2 (BDÜ 2017).[7] This probably reflects professional practice, at least for translation between languages of major diffusion: translators who work into their L2 also tend to work into their L1; in other words, they are bidirectional translators.

The wide demand for translations into English in Europe is also due to the fact that English is currently the European Union's most important working language. In 2013, the Directorate-General for Translation, the European Commission's translation service, translated 250,000 pages into English (European Commission 2014, 8), almost twice as many as ten years previously and 60 % more than the runner-up, which was the number of pages translated into German (Directorate-General for Translation 2009, 3). One of the reasons for this sharp increase between 2003 and 2013 was

5 A language of limited diffusion is spoken only by a limited number of people as L1 and is rarely taught and therefore acquired as an L2 outside its area of diffusion.

6 Here, I only listed examples of countries in which translators work from those countries' official language(s) into an L2. For L1 speakers of Japanese who translate into and out of English in the UK, see e. g. Sakamoto (2017).

7 The search in BDÜ's member database allows to discriminate between L1 and L2 speakers of the source or target languages. For my calculation, I deducted entries from translators who stated that they would not accept any commissions. The absolute numbers are 2410 for E-G and 1719 for G-E, as of 21/09/2017.

that in the European Union with its 24 official languages, English is one of the three procedural languages, together with German and French. This means that every document within the European Commission has to be available in at least these languages in order to have it discussed. Likewise, incoming documents from the member states are often translated into English for information purposes (European Commission 2010, 75). Another reason is that English is also used as a relay or bridge language for texts that have to be produced in all official languages. This means that because there are not enough translators to cover all the possible language versions, some translators use the English translations as the source for their own translations. Nevertheless, someone has to produce all these English translations in the first place. Up until 2009, the European Commission only recruited translators who worked into their L1 as this was considered the gold standard and there were enough translators available for the official languages of the European Union at the time. However, in light of the enlargements that have added another 13 official languages since 2004, the European Union had to find another solution. They introduced what they called *two-way translation* as an official practice (European Commission 2010, 48). This means that translators at the Directorate-General for Translation with much sought-after language combinations now also translate into their L2 in order to make good use of their skills. As, apparently, even Brexit will not have a great impact on the European Union's high demand for translations into English (Carey 2016), translators for the European Commission will continue translating into and out of their first languages, at least in certain language pairs. What may seem like an obvious solution to the European Commission's problem was in fact a great step towards accepting a practice that has existed for a long time in the European translation market: translation into L2, which means to a large extent into English.

Not only is English the most important working language within Europe at present, it is also the global lingua franca. One estimation assumes that L2 speakers of English outnumber L1 speakers of English by two to one (Simons and Fennig 2018). The use of *English as a lingua franca* (ELF) has also increased the demand for written communication in English. Whereas in some situations, texts may be produced directly in English when the people do not share a first language, in other situations texts need to be translated into English for a target audience that probably includes a high proportion of ELF users. It would therefore make a lot of sense for translation also to be performed by L2 speakers of English, who have had the experience of ELF situations and embrace the necessary skills and competence as ELF writers.

Having introduced the topic and my general research approach above, I present the research questions forming the basis of the investigation in

Chapter 2. Chapter 3 then situates these research questions in a theoretical framework based on earlier research on translation into and out of a translator's first language, on the cognitive effort translators exert and on combinations thereof. In Chapter 4, the overall research design and details on the two setups for data collection are explained. Chapter 5 reports on the results for the translation process activities and draws comparisons between the groups of participants. Chapter 6 does the same for the translation products. In Chapter 7, the results for the measured indicators of translation effort are presented and compared across and within groups. Translation effort is then correlated with professional experience, productivity and product quality in Chapter 8. The study's findings are discussed and situated within the discipline in Chapter 9, which also discusses the study's impact and relevance and sketches further research paths as well as pragmatic applications.

2 Research questions and scope

This study uses an empirical approach to investigate professional non-literary L2 translation by analysing translators' processes, products and effort. It combines various data elicitation methods as well as results from the three units of interest. It meets criteria of rigour for empirical research in translation studies and considers methodological issues (as suggested by e. g. Gile 2016 or Neunzig 2002). Its distinguishing characteristic is its commitment to construct validity: Firstly, including experienced professional L2 translators as participants ensures that the goal of gaining insights into professional L2 translation can be achieved. As translation into a second language is usually performed by translators who also translate into their first language, bidirectional translators are my main group of participants. In order to compare their translation behaviour with that of translators who work in the same language pairs but only into their first language, unidirectional translators are also included. Secondly, keeping the experimental situation as naturalistic as possible and collecting data as non-invasively as possible promotes ecological validity. Finally, using a mixed-method approach facilitates data triangulation and thus supports investigating various aspects of directionality from several perspectives. This allows for a more complete insight into the complex topic and adds to the validity of my research project. My study balances comparability of results and ecological validity by having the participants translate in a laboratory setting but letting them work in the language combinations and translation directions that they usually do and proceed as they normally would.

The study addresses the following three research questions:

1. In what ways are the processes of professional L2 translators similar to or different from the processes of professional L1 translators?
2. In what ways are the products of professional L2 translators similar to or different from the products of professional L1 translators?
3. In what ways is the effort for translation of professional L2 translators similar to or different from the effort for translation of professional L1 translators?

All three questions include the comparison of central features of translation with regard to whether the translators work into their L2 or L1. Therefore, they also inherently address directionality, that is whether differences

between those features are potentially related to translation direction. In order to delimit the scope of my study, the key terms are defined below.[8]

A *translator* is a person who produces a written text for an intended audience in a future communication situation. She[9] writes that *target text* (TT) in a *target language* (TL) and uses another text, the *source text* (ST), which is written in another language, the *source language* (SL), as a main source of information. Her task, which is called *translation*, is thus a complex cognitive activity involving at least two languages.

A *professional translator* is a translator who is paid for her work and who adheres to professional standards with regard to service delivery and ethics (Pym 2011; Jääskeläinen Kujamäki and Mäkisalo 2011). Moreover, a professional translator has formal translation training or accreditation and translates non-literary texts.

A *translation process* (TP) starts when the translator accesses a ST and ends when she completes the respective TT.[10] A TP is performed at a computer, manifests itself in observable activities and results in a TT. When performing a TP, the translator expends cognitive effort, which I refer to as *translator effort*.

With respect to the level of analysis, I am guided by the principle that the "granularity we adopt may lead us to interpret an observed component as an indicator of one construct or another: the very notion of what the task is or entails has a direct influence on data interpretation" (Muñoz and Cardona 2018, 4). As I want to investigate whether directionality has an effect on translators' processes, products and effort, I use the translation process as unit of analysis and then divide it into its main observable and countable activities. Based on these activities, I then suggest an array of indirect indicators of translator effort, which in combination provide information about the overall effort for translation. With regard to assessing product quality, I use the target text as the unit of analysis. The results of product quality and translator effort are then correlated with participant data to add to the understanding about the effects of directionality.

As for the language combination, I opt for German and English since they are two major languages in Europe and translation between them is

8 I acknowledge that my definitions of professional translation and my presuppositions about translation in general are Eurocentric. For an enlightening call to reconceptualise Western ideas about translation see Tymoczko (2006).

9 I use the feminine gender for translator throughout my study. This facilitates formulation and reflects the predominance of women in the translation profession worldwide (Pym, Grin, Sfreddo and Chan 2013, 78).

10 For definitions of the translation process in a wider sense see e.g. Schubert (2013, 151), Heidrich (2016, 223–224) or Risku, Rogl and Pein-Weber (2016, 8).

therefore in demand. As I am interested in professional translation and as I want to study translators in action, it makes sense to limit my catchment area to the German-speaking part of Switzerland and adjoining regions. Translators into English probably produce texts that will be used by ELF readers as often as by L1 readers of English. Therefore, it may make sense to see translation into English as ELF communication and integrating specific communication strategies may become opportune. It is as yet unknown whether translators translating into English already consider or even apply such strategies. Exploring this is beyond the scope of my study but it is important to a least raise this point here. In the light of the high demand for translations into English and the lack of recognition of L2 translation practice, it is very likely that a lot of the work is performed by translators who do not have formal training or accreditation as translators. However, including them in my investigation would have weakened the focus on professional practice.

The next chapter situates my study within Cognitive Translation Studies and presents concepts and empirical research relevant to answering my research questions.

3 Theoretical underpinnings and literature review

This chapter presents my research framework within Cognitive Translation Studies, introduces the concepts of directionality and translator effort and reviews how they have been investigated in Cognitive Translation Studies so far. As this study aims to investigate directionality and translator effort by analysing the actual translation processes and products, they also form part of the review of previous research relevant to my study. Furthermore, this chapter discusses different approaches to translation product evaluation suitable for directionality studies and gives a short overview on directionality research in Interpreting Studies. It concludes with a summary of the literature review and implications for my study.

As I want to empirically investigate the real-world phenomenon of professional L2 translation practice, my study follows in the research tradition of *pragmatism*. Pragmatism, as for example defined by Phakiti and Paltridge (2015, 17), bridges the original divide between the qualitative and quantitative research paradigms as represented by positivism and constructivism and thereby contributes to knowledge. As I investigate professional L2 translation by analysing the translation processes, products and effort for translation of translators performing translation tasks in different translation directions, my research is situated in the discipline of Translation Studies. Since I define translation as a complex cognitive activity and I aim to determine the amount of cognitive effort involved in its performance, my approach fits within a subfield of Translation Studies called *Cognitive Translation Studies* (CTS) that, as suggested by Halverson (2010, 349), investigates the cognitive processes of translators (and interpreters) irrespective of the underlying cognitive paradigm. CTS can be divided into two substrands: *translation process research* and *cognitive translatology*, which I adopt as my research framework. While the two frameworks share methods of data elicitation and analysis (see e. g. Krings 2005, 347–356), they differ on their view of translation. Translation process research has mostly been informed by computational or cognitivist views of cognition, albeit often implicitly, whereas cognitive translatology explicitly draws on the paradigm of situated and embodied cognition[11] (Muñoz 2017, 567). I will therefore use CTS when I discuss matters that concern

11 Acknowledging that the concept is now also called 4EA cognition for embodied, extended, enacted, embedded and affective, I use situated and embodied cognition based on Risku (2016, 92), who was among those researchers who introduced it to Cognitive Translation Studies.

the subfield as such and cognitive translatology when I refer to my study's framework in particular.

From its start, CTS has been interested in the cognitive processes underlying the translator's linguistic decisions (Krings 2005, 344). It has strived to achieve insights by a systematic investigation of how target texts come into being, that is by observing translators' activities and making inferences about their mental processes (Hansen 2003; Alves and Gonçalves 2013; Livbjerg and Mees 2003). This is possible as it is one of CTS's basic assumptions that cognitive activity has behavioural correlates that can be observed and measured (Jakobsen 2014, 75). Before the opportunities to gather information about the translation process became available, comparison of the target text(s) to the source text had been used to make inferences about translators' problems, decisions and strategies. However, this approach was guided by what are now recognised to be misconceptions: it assumed that the translation task is a linear cognitive process and that each target text is the product of exclusively conscious decisions on the part of the translator (Martín 2008, 12). It often involved pure guesswork as it ignored what Krings (2005, 348) calls the *process-product ambiguity*: Identical target texts (or parts thereof) can be the result of identical or different processes and – in the opposite direction – different processes can lead to identical or different target texts. CTS has acknowledged this ambiguity by investigating cognitive processes in translation via an analysis of the translation process. Among others, the following complex or higher-order cognitive processes have been associated with translation: problem-solving as manifested in e. g. pause behaviour (Dragsted 2005; PACTE 2005; Couto-Vale 2017), decision-making as manifested in e. g. concurrent verbalisations during information searches (B. Nord 2002), evaluating and monitoring written output as manifested e. g. in target text reading (Jakobsen and Jensen 2008), creativity (Bayer-Hohenwarter 2009), metacognition (Shreve 2009), or considering needs of the potential target audience (Apfelthaler 2014). Less complex or lower-order cognitive activities include handling computer peripherals (e. g. typing[12]), switching between tasks or navigating within a program. For both types of processes, it is assumed that they use fewer cognitive resources if they are automated, which in term leaves more resources for demanding activities that emerge during task performance.[13] However, psychological reality suggests that it is often not clear where lower-order or unconscious

12 The amount of cognitive resources or attention needed for typing depends on the level of skills. Touch-typists seem to be able to activate high-level processes in parallel to typing (Olive and Cislaru 2015).

13 A situational amount of automated processing interspersed with conscious control and a high metacognitive ability have been associated with expertise and expert behaviour (Shreve 2006).

processes end and higher-order or conscious ones begin, so these processes may be better considered as the two poles of a continuum between which most of human mental activity happens (Muñoz 2014, 24).

Methodological and technological developments, with parallels in other areas of applied linguistics such as writing process research or in other disciplines such as psycholinguistics, have expanded CTS's toolbox of methods (O'Brien 2013). It now contains methods to elicit verbal data during or after the translation task, to log the use of computer peripherals (referred to as keystroke logging), record screen activities, capture eye gaze patterns or changes in pupil size (referred to as eye tracking), or measure brain functions (Krings 2005, 348). With the growing awareness that every method has its drawbacks and blind spots, multi-method approaches have become good practice in the field. With the combination of various techniques, data can then be triangulated in order to minimise the distorting effects of a single method and to allow for more precise analysis and a certain degree of cross-verification of results (cf. Krings 2005 or Ehrensberger-Dow and Massey 2008). All this research effort and rigour often referred to by the term translation process research, is aimed at providing an ever more robust picture of all the mental processes involved in the translation process.

The strand of CTS that this study draws on is cognitive translatology[14], which defines translation as a situated and interpersonal activity rather than an intercultural or interlinguistic one (Muñoz 2010b, 154). Indeed, translation had been defined as a situated activity for quite some time, initiated by functionalism, which placed the translator at the centre of attention and embedded her in a series of activities and actors such as client, commissioner, colleagues, and target audience. From the perspective of situated and embodied cognition, however, the cognitive act of translation extends to the body of the translator herself, to her task-based interactions with artefacts and with individuals of her network (Risku 2016, 78–86). Ultimately, situated and embodied cognition stipulates that "[c]ognitive processes are the consequence of interacting with the environment and they affect the environment" (Muñoz 2016, 155–156). As such, human cognition can be perceived as an architect who creates and collaborates rather than an archaeologist who digs for memories and knowledge buried in her brain.[15] Nevertheless, the units of analysis in cognitive translatology have ranged from micro pauses in TT production (Muñoz and Martín 2018) to societal expectations on translatorial decisions (Risku, Rogl and Pein-Weber 2016).

14 For an introduction to cognitive translatology see Muñoz (2010b).

15 My interpretation of Risku (2002, 66).

3.1 The concept of directionality in translation

Terminology to describe and discuss directionality varies according to the context in which it is used. In translation training, a student's languages are named A, B or C with A usually being her *native* language (sometimes also called *language of habitual use*), B being her first active *foreign* language, and C being her first or second passive foreign language. A translator typically qualifies in the language combination A-B-C or A-C-C. With an A-B-C combination, she has the translation versions A-B, B-A and C-A, which includes translation into the second language. With A-C-C there are only the C-A versions, that is translation into the first language. Of course, these categories also depend on what languages the training institution offers. It may well be the case that a bilingual student has to elect her weaker language as her A language and cannot even have her stronger one as one of her working languages. In this respect, life as a professional is less restricted and translators may choose additional language versions and translation directions. As this study is interested in professional practice, it seems reasonable not to use the terminology of the educational setting to describe translation direction. To use *translation into a foreign language* instead, does not seem to be an option, either, as I agree with Risku that a professional translator's *foreign* languages are not really foreign to her (Risku 1998, 224). In addition, the expressions *native language* and *mother tongue* are not appropriate either, as they include the idea of there being a rightful owner of a language. The notion of leaving a zone of safety also reverberates in the German *Hin-Übersetzung*[16] for translating into a foreign language, whereas *Her-Übersetzung* describes the default direction. In French, the term is *le thème*, which was originally used in the context of word-for-word translation exercises into Ancient Greek and Latin and therefore has a connotation of being rare and superfluous. The European Union replaced its original term *two-way translation*, which described the practice of bidirectional translation perfectly, by *reverse translation*. Reverse translation seems an unfortunate choice as it can easily be confounded with back-translation, which is trying to reconstruct the original text from its translation. The Spanish *traducción inversa* and the English *inverse translation* suggest that the translator is going the wrong way, and according to Pym (2011, 84) their use is therefore "ideologically loaded and professionally indefensible". I agree and thus call this translation direction translation into a second language or *L2 translation* (although it may also be L3, L4 etc.) as compared to *L1 translation* (while L1 could refer to two languages in the case of bilinguals). The terms *L1* and *L2* are adopted from the field of second

16 A word-for-word translation for Hin-Übersetzung is *translation to there*, whereas for Her-Übersetzung it is *translation to here*.

language acquisition and commonly reflect the chronology of acquisition. By using these terms, I hope to avoid the negative connotations that some of the other expressions have.

L2 translation has traditionally been practised and may have a negative connotation in some parts of the Western world. In China, literary as well as non-literary translation has a tradition of being done into the L2, often in collaboration with an L1 speaker of the target language (Hung 2005; André 2010). Teams of L1 and L2 translators have also been frequent and successful in Bible translation (Pokorn 2000, 67) and have been posited as its basic principle (Nida 1964, 153–154). Russia has a long tradition of L2 interpreting as, apparently, the commissioners of interpreting jobs favoured trusting the person who was an L1 speaker of their own language. This principle may also have been enacted in the former Soviet Union's world translation project (Tymoczko 2018, 158). L2 translation has also played an important role in the Arab-speaking world in its effort to export cultural goods (Jamoussi 2015). Different notions of translation have been replaced by a mystification of the supposedly perfect bilingual and solitary translator in the Western tradition. Its origins can be traced to the rise of nation states in Europe (Pokorn 2000, 62). With it, the concept of national languages or one national language became an important means of demarcating one nation from others and stressing the former's uniqueness (Pokorn 2000, 62). Every citizen would then also *inherit* that official language and become an owner of it. This development seems to have fostered the notion of the supremacy of the language one *was born with*, also referred to as one's native language or mother tongue. This notion of ownership was transferred to the translator who was declared owner of the target language (Pokorn 2000, 62). The fact that many early published translation scholars were speakers of major languages may have added to the widespread dissemination of this exalted view of the translator (Tymoczko 2006, 17). For communities with languages of limited diffusion, the practice of L2 translation seems to have been tolerated as the lesser evil as, otherwise, their cultural and economic exchange would have suffered. However, this is a biased view and as I explained in the introduction, L2 translation is not only practised all over the world but also respected as a professional practice (Whyatt and Kościuczuk 2013). As a consequence, it has to be considered in Translation Studies. This is done with the concept of translation direction or *directionality*, which has been taken into consideration in models of translation to varying degrees.

Models of translation competence generally state the necessity of *bilingual competence* without prescribing the translation direction in which a translator has to work (PACTE 2003; Göpferich 2008). In their model, the PACTE group defines the bilingual subcompetence as the "[p]redominantly procedural knowledge needed to communicate in two languages" (PACTE 2003, 58), which they then specify as pragmatic, socio-linguistic,

textual, grammatical and lexical knowledge. As their study participants had to work into both translation directions, it can be assumed that the PACTE researchers did not ignore directionality but decided that it was not an issue. Göpferich's model (2008, 155–156) entails communicative competence in at least two languages, which is specified as competence in language production for the target languages and competence in language comprehension in both the target and the source languages. At the centre of both PACTE's and Göpferich's models is *strategic* competence, which seems to be responsible for a task-based activation of and collaboration between the other competences. In addition, Göpferich places *motivation* alongside strategic competence, which "which may be both intrinsic (enjoying translating) or extrinsic (payment, fear of compensatory damages, etc.)" (Göpferich 2009, 22). She also stresses the importance of the translator's self-concept and professional ethos as a basis for translation competence (Göpferich 2008, 156). As these two models of translation competence were originally the results of top-bottom hypothesising about cognitive processes and not of empirical observation, the relation between the different components and how they actually interact during task performance is not yet clear and awaits further investigation (Marín 2017, 13). Risku's model of the *cognitive skills* of translators differentiates between beginners and experienced translators based on four categories: macrostrategy formation, information integration, planning and decisions, and self-organisation (Risku 1998, 244). Risku (1998, 244) identifies the main guiding image that beginners have of translation as *signal transmission* whereas experts see translation as *meaning construction* that considers the specific translation situation, of which presumably directionality would be one aspect.

The European Master's in Translation network defines five areas of competence as standards for translator training and translation competence (EMT Board 2017): language and culture, translation, technology, personal and interpersonal, and service provision. Directionality is not mentioned in any of those areas. With regard to language competence, admission to studies at Master's level requires skills in at least two working languages at CEFR[17] level C1, which is the second highest level in that reference system. Upon graduation, "the translator's main target language should be mastered at CEFR level C2" (EMT Board 2017, 6). This explicitly allows translation into the L2, which makes absolute sense for a competence framework applicable on a European level. It also seems reasonable to state a CEFR level of language competence instead of only mentioning native or bilingual competence, which per se does not indicate any level (cf. Kearns 2007, 171–174). It is noteworthy that in comparison to its predecessor of 2009,

17 CEFR: Common European Framework of Reference for Languages (Council of Europe 2018).

the 2017 framework has a separate section on personal and interpersonal competences (EMT Board 2017, 10), which stresses their importance for graduates' adaptability and employability and reflects an appreciation of translation as a professional and not merely a linguistic practice.

The European and the International Standards for *Translation services – Requirements for translation services* – EN 15038 and ISO 17100, respectively – list five categories of professional competences of translators without referring to directionality (EN 2006, ISO 2015). Linguistic competence is simply defined as source language comprehension and target language fluency. More weight is put on the professionals' qualifications: they require a minimum of (a) a translation degree from an institution of higher education or (b) another university degree plus two years of full-time professional experience or (c) five years' experience. As both standards require revision by an external reviser with the same competences as the translator, it seems that many L2 translators seized the opportunity when the EN 15038 was introduced in 2006 and became certified. For many of them, complying with the requirements was easy, as external revision procedures were already in place, maybe in contrast to L1 translators who often do without external revision.

Overall, while these models of translation competence do mention language skills, they do not specify translation direction or make assumptions about appropriate professional behaviour related to directionality. In the case of competences, language skills are not translation specific and can be considered an actual *pre*-requisite to develop translation competence. Kumpulainen (2016, 3) even reduces translation competence to *interlingual text production competence* arguing that this is the only skill that is needed in every translation situation.[18] In the case of establishing standards, it makes sense that professional qualifications influence the delivery of good quality services and not translation direction.

However, there are a limited number of models that are explicitly or implicitly concerned with L2 translation. Based on empirical data mainly from translators with Arabic and English, Campbell (1998) developed an L2 translation competence model consisting of three relatively independent components: textual competence in the TL, monitoring competence that includes text revision and disposition, which relates to an adequate combination of risk-taking and persistence. Campbell's main aim was to portray L2 translation "as a normal human activity supported by the assistance and insights that can be gained from the methodical application of linguistic theory" (Campbell 1998, 176). Campbell suggests also

18 In this respect, Kumpulainen (2016) agrees with Schrijver (2014), who showed that increased text production skills help translation students to acquire translation competence.

considering these competences in quality assessment. He was well ahead of his time with his suggestion that not only evaluating the product but also the process is important.

Kiraly (1995, 101 and 1997, 156) developed a psycholinguistic model of the translation process based on concurrent verbalisations of participants during a translation task. Kiraly then transcribed the recorded verbalisations and used them as think-aloud protocols for his analysis. His participants were students of translation and professional translators who translated from German into English, their L2. Kiraly's model (1997, 149) represented the translator's mental space and its three principal components were (a) sources of information, including long-term memory, source text input, and so-called external resources (paper, electronic or human); (b) the intuitive workspace, where relatively uncontrolled and subconscious processes happened; and (c) the controlled processing centre, where conscious strategies were applied. While Kiraly acknowledged that his participants translated into their L2, he did not elaborate on any potential issues of directionality. He later focused on the importance of the students' self-concept as translators and suggested that this can only develop by project-based authentic learning experiences. While in 2000, Kiraly mentions the benefits of project work for L2 translation training in particular, he does not seem to consider his further investigations and elaborations in the light of directionality (see e.g. Kiraly 2005). Although his own role in the classroom has apparently changed from instructor to facilitator, he is still "the sole native speaker informant in the class" (Kiraly 2005, 1107) with all the repercussions this may have (Pokorn 2009, 205).

Unlike in foreign language teaching, where non-native speaker teachers have been recognised as competent instructors for a while now (Selvi 2014; Aneja 2016), the TL native-speaker principle in L2 translation teaching has proved extremely tenacious, irrespective of its shortcomings (Hagemann 2016). As an L2 translation teacher, Wimmer (2011) has not probed that attitude but recognised the need for student empowerment. In order to address this need, she developed and validated a didactic model of L2 translation that is based on findings by Neunzig and Tanqueiro (2007) and Grauwinkel and Neunzig (2008). They had used the prominent and popular model of the translation process (Jakobsen 2002, 192–193) consisting of three distinct phases called *orientation*, *drafting* and *revision* phase to describe translator behaviour. They had noticed that the professional translators that the PACTE group monitored for their study on translation competence tended to proceed differently depending on directionality. When translating into their L1, the translators apparently made use of their L1 competence during the revision phase by correcting errors or tackling problems they had postponed during the drafting phase. Into their L2, the translators also made use of their high L1 competence by investing

more time in identifying potential difficulties in the ST before starting TT production. This difference in activities was also apparent in the different proportions of time they spent on each of the three translation phases with regard to directionality (Wimmer 2011, 124). In her PhD study, Wimmer (2011, 123) proposed the following two didactic models to explain the respective roles of the L1 and the L2 in translation:

- for translation into L1: ST analysis (in L2), information seeking (in L2 or L1), transfer (in L1) and reformulation of TT (in L1)
- for translation into L2: ST analysis (in L1), reformulation of ST (in L1), information seeking (in L1 or L2), and transfer (in L2).

Wimmer (2011) then validated the model of the L2 translation process by comparing the development of one cohort of students that was introduced to the model to one that was not. She had both groups do a pre- and a post-test. In the post-test, her experimental group actually produced TTs with a higher acceptability – of five predetermined potential translation problems – compared to their pre-test performance, while the comparison group did not make such progress (Wimmer 2011, 219). Although Wimmer's research design is problematic as she had Spanish-German translators as her experimental group and Spanish-English translators as her control group, her model has its merits for didactic purposes. In its approach to incorporate L1 language skills as an additional set of skills for L2 text production, it resembles approaches in developing L2 writing skills (see e. g. Göpferich 2017). Obviously, Wimmer's model is targeted at translation students whose L2 skills do not yet meet those expected from professionals. The students' inability to revise their L2 TT is an indication of insufficient L2 text production skills (see e. g. Leijten and Van Waes 2015). This may also have been the case with the professional translators who participated in PACTE's study, as only about half of them reported that they did translation into the L2 (Kuznik 2017a, 104). As such, Wimmer's model may serve as what Risku calls a *scaffold* that professionals will dispose of by and by (Risku 2016, 96). Therefore, as she herself acknowledged, Wimmer's model cannot be used as a basis for explaining professional L2 translators' behaviour (Wimmer 2011, 220).

For the sake of completeness, Krings' models (1986) are also mentioned here. Although his participants were learners of French rather than translation trainees or translators, he did investigate differences in behaviour in performing translation tasks into L1 and L2. While performing the tasks, the participants verbalised what they were doing. Based on his analysis of those think-aloud protocols, he then modelled a flowchart for each translation direction. He noticed differences with regard to types of strategies mentioned. However, as the models are almost

sequential, they do not seem to adequately represent the complexity of the translation process (TP) with its iterations and backtracks or the behaviour of participants with translation experience. However, Krings' choice of participants and method to collect data was not uncommon at the time. Lörscher (1991) had advanced learners of English as participants and also used thinking-aloud to investigate problem-solving and decision-making in translation. He hypothesised, based on Krings' (1986) findings, that "[s]ubjects are primarily faced with *production problems* when translating *into the foreign language*, and with *reception problems* when translating *from the foreign language*"[19] (Lörscher 1991, 96). This is a strong indication that Lörscher realised that his participants' L2 skills were not good enough for translation purposes. For L2 language learners, translation into the L2 may indeed be more difficult than into L1. Risku (1998, 223–224), for example, argues based on findings in writing process research that difficulties in problem-solving on a low level of abstraction may not only disrupt the production of the current text sequence but also the planning of following sentences. As the translation task already proved cognitively demanding for Krings' and Lörscher's participants, thinking aloud probably added to this cognitive load. In addition, commenting in their L1 probably interfered with the necessity to produce text in their L2. In short, they did not have enough cognitive resources available for both activities of commenting and performing the translation task.

3.2 Directionality, translation processes and products in Cognitive Translation Studies

The explanatory power of investigating potential directionality effects has been increased by including participants with translation experience. In her PhD thesis, Hirci (2009) had twenty students of translation translate under two conditions from Slovene (their L1) into English (their L2). They translated one text with access to paper resources only and one text with access to paper, electronic and online resources. As Hirci wanted to investigate the use of paper and other resources and their effect on productivity and product quality, she collected both process and product data with the methods of keystroke logging, think-aloud protocols, retrospective interviews and questionnaires. Results showed, among other things, that there was no significant difference with regard to speed and efficiency between the two conditions. Hirci (2012) then had the TTs evaluated by nine L1 speakers of English who taught translation or linguistics. The evaluators

19 Lörscher's emphasis.

graded the TTs, probably without having access to the STs, according to three categories *good* (only some minor mistakes), *acceptable* (some revision required) or *inadequate*. Results suggested that access to all types of resources and the use of electronic resources in particular had a positive effect on product quality (Hirci 2012). Consequently, Hirci encouraged the fostering of information seeking skills in L2 translation teaching as a preparation for professional practice. Interestingly enough, Hirci (2009, i) justified her sample of translation students instead of professionals by arguing that the students were up-to-date as to the use of non-paper resources for translation. Since then, things have changed substantially: in a survey by Gough (2015, 132), the vast majority of the more than 500 respondents, all professional translators, reported that they spent at least 80 % of their information-seeking time in online resources as compared to 10 % or less in paper or electronic resources, respectively. Nowadays' easy access to and large range of online resources has probably facilitated the translators' searches for information, which L2 translators are generally said to depend upon.

Enhancing L2 translation training was also the main aim of Pavlović's PhD study (2007a). She had translation students perform translation tasks in groups of three under two different conditions after their final exam: collaborative translation into L2 (Croatian-English) and into L1. In addition, Pavlović used control experiments aimed at comparing collaborative and individual translation. Analysis and categorisation of the groups' think-aloud protocols with regard to decision arguments suggested that "they seemed to have much stricter criteria for monitoring the suitability of particular tentative solutions" in L1 than in L2 translation (2007a, 183). When it came to discussing the ST, though, there was no difference in quantity, which is against the common notion that ST comprehension is more difficult in L1 translation (2007a, 184). Also with respect to types and numbers of translation problems encountered, there was no difference between translation directions (2007a, 181). In addition, the types of information resources used seemed to be more a question of group or individual preference than directionality (139). Pavlović also had the translation products assessed based on their need for revision. Three external evaluators assessed all TTs by marking parts that corresponded to either of the two categories of *must revise* or *revise if there's enough time* (Pavlović 2007a, 76). The evaluators had access to the STs but were not informed about who had produced which TT. They evaluated the TTs in two rounds: first on their own and then together in order to reach consensus. To facilitate quantitative comparison, Pavlović then assigned weighted scores to the two categories and calculated a total score for each text. Results of product quality showed that collaborative tasks produced products of higher quality than individual tasks did (Pavlović 2007a, 186) and

collaborative L1 translation produced higher quality than L2 translation did (Pavlović 2007a, 182). Pavlović suggests using collaborative L2 translation tasks to generate and evaluate more tentative solutions, to boost student's confidence by immediate peer feedback and to enhance the profitable use of information resources. Interestingly, Pavlović (2007a, 157–158) reported that there was no correlation between the translators' ratings of enjoyment, task difficulty and satisfaction with product quality. Pavlović investigated translation students on the cusp of entering professional life. Other studies have included translators with professional experience.

For example, I investigated professional translators' self-concept and its potential relation to directionality (Hunziker Heeb 2016). *Self-concept* can be understood as how the translator sees her roles and responsibilities. I had two groups of translators with German and English as their languages. The first group was used to translating both from German into English and from English into German (i. e. bidirectionals). The second group only translated into their first language, i. e. in only one of the language versions. After they had performed a translation task, they commented on the whole process. Those verbalisations had been cued by a screen recording of the translation process, which included visualised gaze data from eye-tracking records, a combination that had been found to produce information-rich data (Ehrensberger-Dow and Massey 2013). I transcribed, encoded and categorised the commentaries according to their foci of attention. Five categories emerged ranging from "(1) focus on literal transfer of single words and phrases, (2) changes to sentence structure, (3) issues of target text quality and (4) loyalty to the source text to (5) awareness of the intended readership" (Hunziker Heeb 2016, 80). Both groups of translators displayed a comparable meta-awareness of the whole range of foci of attention and the need to balance them during task performance, which I interpreted as robustness of the translators' self-concepts, irrespective of directionality.

Other studies have focused on the relationship between directionality and coping with uncertainty. Lorenzo (1999, 2002b and 2003) explored professional L2 translators' strategies to cope with uncertainty elicited by a non-routine task, which is required by a competent translator. Her participants were twelve experienced Danish-Spanish translators who had no specialisation but who differed with regard to employment situation and age (Lorenzo 2002b, 94). They translated two different STs at their workplaces into their L2, without a time limit but with remuneration. Lorenzo used keystroke logging, thinking-aloud, screen recording and retrospective verbalisation to monitor the translation tasks. Based on her analysis of the think-aloud protocols, Lorenzo (1999, 127 and 2003, 112) found that her participants tended towards *message adjustment* and concentration on keywords in the task they were less familiar with. At first sight, this procedure seems reminiscent of Wimmer's (2011) didactic model that encourages

students to compensate their lack of L2 skills by reformulating the ST in their L1. However, Lorenzo (2003, 94) suggested that her participants consciously chose message adjustment as their strategy of *playing-it-safe* as opposed to *risk-taking* in order to prevent misunderstandings by the target audience. Lorenzo (2002b, 116) suggested that the frequency and purpose of searching for information in L2 translation differed from that of L1 translation due to the L2 translators' particular approach to managing uncertainty. She argues that maybe because of this, some (e. g. Gross 2000) insist that L2 translation only be done with standardized texts, following a pattern of actions, performing repetitious routine tasks. Lorenzo (2002b, 90) criticises this reduced view of translation in general and of the potential of professional competence in particular.

Lorenzo (2002b) also had the translation products evaluated. Seven evaluators (translators, L1 writers and members of the target audience, i. e. website readers in this case) graded the TTs according to the common exam grading system in Spain. The only guidance given was to ask whether the TTs fulfilled their purposes (Lorenzo 2002b, 97). Lorenzo then calculated an average mark. The four translators with the best performance apparently had performed message adjustment and used only monolingual dictionaries and parallel texts as information sources, if at all, whereas the translators with the poorest performance had made frequent use of bilingual dictionaries (Lorenzo 2002b, 107–108). Lorenzo discourages the exclusive use of bilingual dictionaries as they would not provide enough information for L2 translators and would therefore be an inadequate means to cope with uncertainty. An additional aspect of Lorenzo's studies (2002a and 2002b) is translation effort. It seems that many studies on effort are also concerned with directionality. One reason may be the general assumption that there is actually a difference in effort related to directionality. Studies that relate effort to directionality are discussed in Section 3.5 on translator effort.

In their investigation into translation competence, the PACTE group (Hurtado 2017) analysed their participants' translation processes and products. They had 35 professional translators and 24 foreign language teachers as participants. To qualify, the professional translators had to be between 30 and 55 years old, have a minimum of five years of experience, work in more than one subject and generate at least 70 % of their annual income from translation.[20] These conditions were based on self-declaration (PACTE 2009, 209–210). A degree in translation was not a prerequisite for participation (Kuznik 2017a, 101). Their L1 were either Spanish and/or Catalan and their L2 was either English, French or German (PACTE 2017b, 95). The participants had to translate a news text into Spanish and

20 I only report on the translators here.

an excerpt of a Spanish tourist brochure into their L2 (PACTE 2009, 210). In order to balance the number of language versions, eight translators agreed to translate out of and into a language that was not their L2 but still one of their working languages (Kuznik 2017a, 100). It is noteworthy that while Catalan was treated as an L1 in the recruiting process, it was only offered as a target language and not as a source language in the translation tasks (PACTE 2017b, 91). This means that for the L2 translation task of a Spanish ST, those ten translators who had reported Catalan to be their only L1 actually translated from their L2 into their L3 or Ln. In total, less than 50 % of the translators reported that they engaged in L2 translation (Kuznik 2017a, 104). The PACTE group found that the translators performed more information searches when translating into their L2 than into their L1 (PACTE 2017a, 292). PACTE concluded with their focus on translation competence: "This would indicate that shortcomings in linguistic and extralinguistic sub-competences, since subjects are translating into a language which is not their own, is compensated by instrumental competence" (PACTE 2009, 227). However, other factors than potential shortcomings may have accounted for this difference, such as the translators' lack of experience in L2 translation or PACTE's methodology to count only those information searches that were related to their five predetermined potential translation problems in the source texts (rich points). In L2 translation, PACTE found a correlation between the time used for those information searches and product quality: The longer the duration, the higher was the target texts' acceptability, which was also determined based on the rich points only (PACTE 2017a, 292). The same positive direct relation was also found between the number of different resources used and TT acceptability (Kuznik 2017b, 227). These relations were not found for L1 translation. Therefore, it seems that the translators made good use of online resources when translating into their L2 whereas into their L1, the efficiency was not that obvious.

PACTE (2009) also evaluated the quality of the translation products. They equalled quality with acceptability, which they defined "in terms of whether or not the solution found effectively communicates (a) the meaning of the source text; (b) the function of the translation (within the context of the translation brief, the readers' expectations, genre conventions in the target culture); and (c) makes use of appropriate language" (PACTE 2009, 217). PACTE applied those criteria to five segments in the two STs they had previously identified as rich points. They argued that the use of rich points offered methodological advantages such as facilitating data triangulation from multiple sources and economical data analysis (PACTE 2009, 213). PACTE (2011, 326) claimed that those five segments represented prototypical problem types or combinations thereof and were the most frequent ones encountered in professional translation. However, their view was not

shared by all participants, who identified different, individual problems as prototypical (PACTE 2011, 333). For each rich point, a description of its content and of its expected TT solution was formulated in advance. Then, the translators' solutions to those ST segments were assessed by two of the researchers and categorised as either acceptable, semi-acceptable or non-acceptable according to the three criteria cited above. In the case of disagreement between the evaluators, other members of the research group were consulted (PACTE 2011, 324). The results were then transformed into weighted scores and its mean calculated. PACTE found that the translators produced much higher product quality in L1 translation than in L2 translation (PACTE 2009, 219). In my opinion, this is not surprising considering that only about 50 % of the translators indicated that they did L2 translation at all (Kuznik 2017a, 104). PACTE reported that only five of the total of 35 translators stated that they actually translated into one of the three languages that PACTE offered as target languages in L2 translation (Kuznik 2017a, 104). It speaks for the confidence of the translators that they agreed to perform L2 translation in PACTE's study at all.

It needs to be stressed that PACTE's research project focussed on the modelling and empirical validation of translation competence and not on L2 translation performed by experienced professionals. All their participating translators had to perform L1 and L2 translation tasks irrespective of the fact that less than 50 % actually had experience in L2 translation. Moreover, the text types in the two translation tasks differed, which also could have had an effect on results (PACTE 2017a, 290). Therefore, as PACTE admits, their findings on directionality should not be over-interpreted (PACTE 2017a, 290). Even considering this, they still present an important contribution to the study of L2 translation and directionality as they indicate that features of information seeking may be related to directionality. By choosing information seeking as unit of analysis, they reinforced the notion that it is an integral part of the translation task (also see Schubert 2007, 96).

In some of the studies discussed above, product quality has also been assessed. In order to be able to situate my approach to TT evaluation, the different approaches to evaluating quality are grouped and discussed in the following section. To my knowledge, though, there appear to be hardly any empirical studies exclusively devoted to the translation product with regard to directionality. It seems that the topic has been investigated from either a process perspective, with or without some product analysis, or a didactic perspective, as can be seen for example in the proceedings of only two conferences exclusively devoted to directionality so far (Grosman, Kadrić, Kovačič and Snell-Hornby 2000; Kelly, Martin, Nobs, Sánchez and Way 2003).

3.3 Evaluation of translation product quality for directionality studies

The product evaluations performed in the studies discussed in the previous section were obviously used in a research context. However, they already exemplify a wide range of possible procedures:

- a mixed group of seven evaluators graded the TTs based on their own criteria using a common exam grading system (Lorenzo 2002b)
- translation and language teachers assessed the entire TTs on a three-point scale (Hirci 2012)
- teachers of L2 translation assessed the level of acceptability of predetermined items on a three-point scale (Wimmer 2011)
- two members of the research group assessed the level of acceptability of predetermined items on a three-point scale (PACTE 2009)
- three bidirectional translators/teachers marked TT items with either revision required or revision suggested, to which weighted scores were then assigned (Pavlović 2007a)
- translation and language teachers with the same L1 as the translators used a common scoring scheme (V. Chang 2011)
- the effort of proof-readers was measured and used as an indicator of target text quality (Whyatt, Kościuczuk and Turski 2017)

Each of these approaches uses a different combination of type of evaluator and evaluation criteria. In general, the choice of combination is guided by the purpose of the evaluation. For example, it can make a difference whether the evaluation serves didactic purposes such as assessing student performance, economic purposes such as quality assurance in a language services provider or research purposes such as evaluating a new tool. Certainly, a set of evaluation criteria is not reserved for one purpose only but can serve different purposes. In addition, if product quality is the main subject of an empirical study, a different approach may be required than if it is a sub-topic. It is probably also the amount of work needed to perform and analyse an assessment that has discouraged CTS researchers from also investigating product quality and correlate it with process activities. Approaches to product quality assessment can be categorised in various ways. In the following paragraphs I first discuss the categorisation by evaluation and then by type of rater. I focus my selection of approaches on their feasibility for my investigation into directionality.

The most widespread and common approach is the marking of errors. It generally involves a predetermined catalogue of criteria – either the translation commissioner's or the evaluator's. While this may help control

for a certain degree of objectivity (Bittner 2014), it may also increase task complexity considerably if the evaluation criteria are complex or foreign to the reviser. There is a risk that error-based approaches are based on phenomena that are *expected* to emerge in translations rather than on *existing* phenomena (Klimant 2016). Moreover, even a fine-grained scheme does not prevent an evaluator from not being able to unambiguously categorise a TT phenomenon. Some evaluation schemes include the possibility of awarding bonus points as for example for creative solutions (Bayer-Hohenwarter 2011). However, Klimant (2016, 205) points to the fact that bonuses are prone to have the same limitations as errors as they may depend on the evaluator's perspective and add an additional level of complexity to her task. Colina (2009, 239) requires that a "general comprehensive approach to evaluation may need to address multiple components of quality simultaneously" and has to be applicable in a professional as well as an educational setting. She proposes a set of descriptive statements that are related to four different components of translation product quality: linguistic form, functional adequacy, non-specialised content and specialised content (Colina 2008, 104–105). Every component is described with four statements about the translation product under scrutiny. The evaluator chooses the descriptor she considers most appropriate – also in relation to the stated purpose and target audience of the text. To determine the total score, each component is weighed in relation to its importance in the particular task and each descriptor is pre-assigned a numerical value. While this approach diverges from the application of an error classification scheme it still involves assessing the TTs with regard to the ST (Colina 2008, 105) and therefore requires language or translation experts or task-trained bilinguals, as used in Colina (2008), rather than laypeople.

Another approach to product evaluation is only considering some parts of the target texts. The PACTE group has used a set of predetermined potential translation problems in the STs as the basis for evaluation. Those rich points represent what PACTE called prototypical translation problems (PACTE 2011, 327). Apparently, they chose this approach in order to be able to handle their elaborate data collection and analysis as they had STs in three different languages for translation into L1 and one ST for L2 translation (PACTE 2009, 213). A disadvantage of this approach is that the evaluators tend to anticipate a certain type of problem-solving behaviour on the part of the translator as well as a certain type of solution. Yet another approach to evaluation is a holistic assessment of the TT. In order to apply this approach, Grabowski, Becker-Mrotzek, Knopp and Weinzierl (2014, 155) suggest so-called naive global measures, that is criteria which even untrained raters can apply because they are simple and global. In their example of assessing text quality in general, they opted for dichotomous judgements such as high/low for a selection of criteria such as the degree of

reader-orientedness. Studies focusing on translation quality often seem to apply a combination of approaches.

In her PhD study on translators as text producers, Schrijver (2014, 100) also used two different evaluation procedures. A professional translator and the researcher herself first applied a holistic score (from 0 to 10) and then an analytic score using an error-based scheme with the following four categories: content (5 criteria), style (4 criteria), genre conventions (4 criteria), and grammatical and orthographical correctness (3 criteria) (Schrijver 2014, 136). Those error categories were weighted according to "how they influence the skopos of translation. Errors that distorted intelligibility of the TT in relation to the brief and the communicative situation were weighted 2, other 1" (Schrijver 2014, 100). However, this approach seems rather time-consuming and laborious. Moreover, the weighting of error types may be influenced by ideology such as the *Skopos* approach applied by the researcher-raters (see e. g. Göpferich, Bayer-Hohenwarter, Prassl and Stadlober 2011, 73), which may not be shared by the text producers. Klimant (2016) also combined a holistic approach with a more detailed one, however applied by two different groups of raters. While the holistic evaluation was performed by representatives of the intended text users, the phenomenon-based one was performed by professional evaluators. They only had to identify and mark the target text segments that they considered in need for revision. Those phenomena were then categorized and counted to determine product quality. A central requirement of the study was that at least three evaluators perform the evaluation in order to increase intersubjectivity and avoid a draw (Klimant 2016, 213). While this principle should be considered in research, it would not be easy to implement in a didactic setting. As an alternative, McAlester (2000a and 2000b) suggested to use editing time as a criterion. He specifically suggested this for L2 translation for didactic purposes, where one evaluator usually assesses several TTs based on the same ST. His main goal was to substitute error-based approaches as for example House's (1997), which he criticised as being too focused on comparing ST to TT on a micro-textual level. He therefore favoured C. Nord's (1991) model, which also considered macro-level requirements such as the intended communicative purpose of the translation and suggested a hierarchy of errors. Nevertheless, he suggested holistic approaches that are less laborious and offer levels of adequacy, thereby reflecting the requirements of professional translation. As he himself realised that his criterion to "correct the translation into a form suitable for its purpose" (McAlester 2000a, 237) left too much room for interpretation, he suggested adding guidelines to prevent evaluators from applying their personal linguistic and translatorial preferences. While his criticism of the complexity and impracticality of the selected models for translation evaluation seemed

justified, he had to acknowledge that his method was not without challenges either, for example with regard to how to actually define the criterion that distinguishes between an acceptable and an unacceptable translation (McAlester 2000b, 138).

Kruger and Kruger (2017) also report on holistic but indirect approaches to translation quality. They suggest measuring the cognitive effort of readers while processing translations. This is similar to Whyatt, Kościuczuk and Turski's (2017) approach to measuring cognitive effort of proofreaders while actually proofreading the target texts. This allows for relative comparisons of translation quality. However, in the case of Whyatt, Kościuczuk and Turski (2017) who investigate directionality, it remains to be seen how they actually compare the different raters' efforts and relate it to product quality and directionality.

It can be assumed that the choice of raters should also be driven by the purpose and goal of the evaluation and of the specific evaluation task (Colina 2009, 255). The commissioner of the evaluation has to decide between experienced or unexperienced raters, professional translation revisers with or without access to the source text or professional editors without access to the source text, task-trained or untrained raters as represented by potential end users. The commissioner's notion of quality also decides on whether access to the ST is considered essential or not. Often, it is the researcher herself who applies her own list of error criteria. While this approach is tempting, as it does not depend on having to sample and instruct raters, it may be difficult to assume impartiality and is therefore open to criticism. As an additional requirement of good practice, evaluation should be performed by more than one person.

With regard to my study, an option would be to have translation or text production teachers evaluate the texts, as they are the ones who are used to assessing a series of texts written by a whole cohort, which professional revisers may not necessarily be. However, professional revisers seem to be more appropriate for assessing the work of professional translators, as this may be a normal procedure in professional life. Then again, revisers are usually not trained to apply the evaluation criteria researchers may want as they generally only need to correct the translations (Doherty 2017, 141). With that in mind and since product quality is only one of several topics of my study and the effort to perform the evaluation and analyse its results needs to be reasonable for a single researcher, I opt for a user-centred approach in which potential members of the target audience, that is newsreaders, assess the target texts. Just as they would as readers of any translated news text, they do not have access to the ST and use their own evaluation criteria. As a consequence, they do not need any pre-task training. In addition, they will only rank and not mark the texts, which further reduces task complexity. While it seems complex to incorporate user feedback already into the actual

translation process as suggested by Suojanen, Koskinen and Tuominen (2015), it seems reasonable to use it for evaluation purposes in research. Admittedly, the approach to translation quality in my study is a narrow one as it only considers the quality of the translation product. Jääskeläinen (2016) and Abdallah (2010), for example, extend translation quality to a multidimensional concept incorporating aspects of product, process and social quality. Jääskeläinen (2016, 89) mentions that considering and investigating these three dimensions of quality is in line with understanding translation as a situated and embodied activity. To incorporate this view, some training institutions have started to combine product with process evaluation to assess their students' translation performance (Massey and Ehrensberger-Dow 2012).

3.4 The concept of translator effort

This section introduces the concept of translator effort, establishes its relation to cognitive effort and cognitive load and examines how they have been investigated with regard to translation, mostly in Cognitive Translation Studies (CTS). It also discusses appropriate measures of translator effort for the purposes of my study.

As with performing any task, translating involves effort. The translator sets up her computer workplace, (re)consults the translation brief, accesses the source text, produces the TT, monitors what she has written, rephrases it, consults the ST, searches for information in online resources and so on. All these observable activities – including keyboard handling and program navigation – are the manifestation of underlying, unobservable cognitive processes of various levels. The total effort that the translator expends during the translation task is what I call *translator effort*. The target text (TT) is then the product of this translator effort. The translator expends this effort in order to get the job done. Otherwise, she would have exerted her effort in vain. Thus, task abortion is only considered as ultima ratio.

The translator herself decides how much effort she wants or needs to expend. She reconsiders her decision consciously or automatically during task performance e. g. because she progresses faster than predicted or she was interrupted or task priorities have changed. Thus, the amount of effort a translator expends on a task is individual. As a consequence, we can assume and we also know from experience that the same task may be performed with low effort by one translator while it is performed with high effort by another (Paas, Tuovinen, Tabbers and Van Gerven 2003). Moreover, it can be assumed that the same task would be performed differently by the same translator on different occasions, e. g. because of differences in motivation or experience. Translator effort therefore

features interpersonal as well as intrapersonal variability. The relation between translator effort, cognitive effort and cognitive load is, however, not completely clear (Jääskeläinen 2010, 220–221).

Performing a translation task uses many cognitive resources or in other words produces cognitive load. Coping with this load and efficiently allocating the restricted amount of resources during task performance involves effort. Every activity during task performance is assumed to be more or less closely related to that task. This assumption also applies to the underlying cognitive processes, especially in cognitively demanding tasks that imply high levels of concentration and attention. The process activities are therefore related to the cognitive processes in the sense that the number and types of activities performed represent the amount of effort invested. The performed cognitive processes manifest themselves in process activities and, conversely, the analysis of these activities allows us to make inferences about the cognitive effort that the translator has expended. Process measures can therefore be used as indirect indicators of effort and of underlying cognitive processes.

In CTS, approaches also exist that distinguish between different types of effort: cognitive, temporal and technical. I will discuss these further down, but in my understanding every activity performed during a task is related to that task under the premise that the translator's main focus is on task completion. Therefore, even actions such as jotting down a shopping list or looking out the window can be seen as part of the TP as they may serve as recovery and prevention of cognitive overload. In other words, there are no translation-unrelated activities and all cognitive processes are related to translation (for a very different and extremely segmented view of translation see Couto-Vale 2017).

Juggling all the constraints and requirements that comprise the translation task uses resources from the cognitive capacity that is thought to be restricted. This notion of limited cognitive resources is the overarching premise for *cognitive load*, a construct often associated with psychological research on learning and the effects of instruction. *Cognitive load theory* and its empirical implementation as *mental effort* are geared to enhance learning by designing adequate means of instruction (Martin 2014). The less effort needed to process task instructions or documentation, the more cognitive resources there are that can be allocated to the task itself, which is learning. Cognitive load theory differentiates between three types of cognitive load that are additive (see e. g. Paas et al. 2003 or CESE 2017): *extraneous* (or *ineffective*) load, *intrinsic* load and *germane* (or *effective*) load. *Extraneous* load is caused by poorly designed instruction material and techniques and inhibits learning. It can and should be reduced. *Intrinsic* load has a task-based and a learner-based component and cannot be manipulated. *Germane* load is the opposite of extraneous load as it is

caused by well-designed instruction and facilitates learning.[21] Studies to validate this multidimensional construct have mostly been performed in relation to learner units in mathematics and other natural sciences. The main issue that has not been overcome is that the single components can be measured neither separately nor directly (Martin 2014 and Paas et al. 2003, 67). Overall cognitive load can only be estimated by determining the mental effort actually expended during the task. Methods have included mainly self-ratings and performance tests (Paas et al. 2003, 64). Apart from its complexity and lack of operationalisability, the construct neglects important aspects such as (learning) environment, self-concept and affective aspects such as motivation (Martin 2014, 612). This also seems to apply to some constructs of cognitive load in Interpreting Studies, where the different types have been related to *listening and analysis*, *production*, *memory* and *coordination* (Gile 2009 and Seeber 2011, 2013). An efficient allocation of cognitive resources – be it from only one unspecific pool as in Gile's *effort model* or from task-specific pools as in Seeber's *cognitive load model* – enables and enhances task performance. This also applies, of course, to translation, with the following factors as potential sources of load: ST difficulty, adherence to client style guide, change of addressees, restrictions to TT length, unfamiliarity of subject field or text type, ergonomic factors, deadlines etc. Translation direction has also been considered to potentially affect cognitive load. To my knowledge, subtypes of cognitive load have been underspecified in CTS but the concept of overall cognitive load has been adopted in many studies, although often only implicitly.

One framework that has been used to study cognitive load in translation is *ergonomics*. The focus has been on professional translation in workplace studies and on human-computer interaction as for example in the special issue of ILCEA[22] (Lavault-Olléon 2016). Ergonomics (formerly called *human factors*) has the aim of "adapting nonhuman elements in a system to fit the human, in order to enhance comfort and performance" (Ehrensberger-Dow 2017, 334). Therefore, its three domains *organisational*, *physical* and *cognitive* ergonomics potentially have an impact on the effort a translator decides, is willing or needs to exert to overcome obstacles in task performance. From the perspective of organisational ergonomics, this

21 The underlying paradigm of cognition emphasises schema construction and automation, which reduce working memory load and thereby allow effective learning. This emphasis as well as the understanding of working memory as short-term information storage may need reconsidering under the recent paradigm of situated and embodied cognition (Martín 2017, 121 and Risku 2002, 67 and Muñoz 2014, 19).

22 ILCEA: Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie.

involves e. g. whether the definition of task including deadline is clear, the respective roles and responsibilities of everyone involved, the availability of information and client material, the translator's position within an organisation or a network of agents. Topics with regard to physical ergonomics are the office environment, workplace equipment, ambient factors such as noise, lighting or heating. In terms of cognitive ergonomics, relevant issues are customising of tools, touch-typing, and layout on screen. It has been shown that disruptions in the human-computer interaction can cause *cognitive friction*, which impedes translators from working efficiently and detracts motivation (Ehrensberger-Dow and O'Brien 2015). While it is obvious that cognitive ergonomics is related to cognitive processes, organisational and physical ergonomics also affect cognitive resource use, e. g. via increasing or reducing stress. This broad view of potential factors influencing task performance is in line with the perspective of situated and embodied cognition that perceives cognition as being enacted in a body that is in constant interaction with its environment. Krüger (2015) posits that adverse environmental factors as well as a translator's impaired physical or mental disposition have an impact on translation performance even if the translator is experienced and competent. An illustrative example of this was presented in a case study (Ehrensberger-Dow and Hunziker Heeb 2016), where the translator's behaviour changed probably because of mental fatigue and hunger. An international survey on the ergonomics of professional translation (Ehrensberger-Dow et al. 2016) established that the characteristics and the quality of the ergonomic settings are closely related to whether a translator works as a freelancer, an employee in the private sector or in the public sector. Although ergonomics does not replace cognitive translatology as my research framework, it may prove helpful in interpreting my results with respect to translator effort and drawing implications from them.

Investigating cognitive processes during task performance sheds light on how the task performer copes with cognitive load. This can be achieved via measuring the cognitive effort the translator expends and this again can only be measured indirectly. This *double indirectness* may have led some CTS researchers to fail to make a distinction between load and effort. However, this is speculation since despite a respectable number of empirical studies on the topic, definitions are scarce while synonymous use of different terms is abundant. Nevertheless, according to Muñoz (2012a, 179) investigating cognitive effort may prove beneficial for CTS as it focuses on processing constraints and not on cognitive architecture, which makes its investigation viable under different research traditions (Marín 2017, 34).

Measures of cognitive load that have been used in empirical research can be categorised as follows: *Psychophysiological techniques* include measures of heart, brain and eye activity, and galvanic skin response. A drawback

of these methods is their potential invasiveness, which may interfere with task performance. *Subjective measures* include self-rating scales or protocols where the task performers report on their effort. In educational psychology, self-ratings are collected either immediately after the task or with some delay. In research on post-editing, self-reports are collected either immediately after working on a single segment or even prior to a task as predictors. Concurrent and retrospective verbalisations as well as process protocols have all been used in CTS. *Process activities from observation* are mainly used in writing research and CTS. They comprise pause and revision patterns, switches in visual attention between the ST and the TT, revision patterns, fluency in production, gaze patterns etc. *Performance measures* are used post-task in educational psychology to test whether learning has been effective, usually operationalised as error rates. Measures from secondary tasks, such as Stroop or other decision tasks, comprise reaction time, accuracy and error rate. However, they have the drawback that they interrupt the primary task.

There are a number of ways that cognitive effort in translation has been investigated, although usually into the L1. Some of them are similar to those mentioned before for measuring cognitive load. Investigations can be grouped according to their units of analysis from the micro level of the word through to ST/TT phenomena, sentences, process phases or overall task. However, units' categories may overlap. In the following, each of these units is introduced with example studies.

The approach of using words or parts of words as units of analysis for cognitive effort has been borrowed from psycholinguistics, which has investigated translation effort in *translation priming* experiments. Participants who are not necessarily translators or interpreters have to produce oral equivalents in L2 or L1 or even just picture naming and their reaction times are measured. These measures and any differences with regard to directionality are explained by a *switching cost* for accessing the relevant language system (de Groot and Christoffels 2006). One of the main research aims has been to develop or test models of bilingual memory (for an overview on different types of models see Heredia and Brown 2004 or de Groot 1997). While informing theories of bilingual lexical access, the task itself does not have much in common with translating a whole text (Tirkkonen-Condit 2000, viii), but models have been used to interpret results from CTS (see e. g. Ferreira 2014). For my research paradigm, studies on effort that are based on translation tasks as I understand them (i. e. starting with accessing the ST and ending with a complete TT) are more relevant.

Phenomena in the source and target text are the next larger unit of analysis that is usually related to the performance of complete translation tasks. Such phenomena include different types of metaphors as an example of predetermined potential translation problems. The translation of metaphors

is assumed to create different levels of cognitive load, depending on how the translator renders them in the TT. Sjørup (2013) and Förster Hegrenæs (2018) investigated this aspect by using different data elicitation methods and by triangulating process and product data. Sjørup (2013) had professional translators as participants and compared their effort for metaphor comprehension to their effort for metaphor production by measuring gaze activity and production time. Förster Hegrenæs (2018) had translation students of different levels translate the same STs and compared their effort for metaphor translation with regard to the development of translation competence using screen recordings, questionnaires and interviews.

In the same category of units of analysis but a different phenomenon falls the effort expended on using different levels of literalness in the TT. Schaeffer and Carl (2014) used several indicators of effort from gaze activity and production time measures for their investigation. Da Silva and Pagano (2017) used indicators from keystroke logging to investigate whether different levels of metaphoricity in primed ST items correlated with the amount of cognitive effort expended to translate them.

In their research on translation competence, the PACTE group also included their views and findings on translator effort, which they called cognitive involvement (2009). Based on screen recordings and retrospective questionnaires, PACTE determined different sequences of actions that the translators performed when they translated PACTE's predetermined potential translation problems (rich points) in the STs. PACTE (2009, 222) defined three possible sequences of actions by the participants when faced with a rich point that they were aware of: the production of a provisional solution, of a final solution or seeking information in documentary resources. PACTE then classified the types of resources used in information seeking "depending on the degree of subjects' cognitive implication in each" (2009, 222). The spectrum ranged from the adoption of a variant offered in a bilingual dictionary, which was assumed to involve the least cognitive effort, to solutions produced without using any so-called external resources, which PACTE attributed the most effort to. It seems that PACTE's preconceptions on the use of certain online resources might have influenced the outcome of their experiments and conclusions.

Using a different approach, Dragsted (2012) did not predetermine rich points in the ST. Instead, she compared the TTs to the ST and identified ST items that translation students had translated differently compared to items that they had translated the same way. Dragsted (2012) then compared the amount of effort involved in translating each of the items using indicators from gaze measurements and keystroke logging, drawing on the assumption that having to choose from a number of variants renders a translation task more difficult and therefore more effortful, which is based on Choice Network Analysis (Campbell 2000).

All the examples mentioned above involved manipulating the text material in order to include the required number and type of stimuli. From the perspective of CTS, it would make more sense to identify rich points in the translation process itself based on indicators of problem-solving behaviour, as suggested e. g. by Muñoz (2014, 24). A potential issue with assuming pre-determined sources of cognitive load is that the performers of the task do not necessarily react in the way the researcher expects. For example, a translator may not identify the same segments as being particularly complex and as needing extra cognitive resources. In addition, operationalising when the processing of predefined items begins and ends is not without its challenges. Sjørup (2013, 199) acknowledges that "it is difficult to formalise a definition for a metaphor's boundaries" in the translation process. She adds that multiple-word units such as metaphors can pose a problem in eye-tracking experiments that prefer the approach of defining areas of interest or investigation that stay within a single line of text. This highlights that the choice of method has an impact on which units of analysis make sense to investigate. Translator effort on sentence level has mainly been investigated in studies on translating with computer-aided translation (CAT) tools or on post-editing machine translation (MT) output. This is due to CAT and MT tools being based on sentence-by-sentence alignment of ST and TT, at least up to now. To use sentences as units of analysis therefore implies that a complete and one to one alignment is actually possible. O'Brien (2006a) used pupillometry to analyse translators' processing of different types of so-called matches from CAT tools while Teixeira (2014) used indicators from keystroke logging (the number of keystrokes) and interviews to compare the effort expended on suggestions from either Translation Memory or MT.

However, measuring cognitive effort on the sentence level seems to ignore some aspects, as post-editors (and translators) also work on larger text or task units. Vieira (2016b, 46) points to this necessary constraint in order to increase the likelihood of reliable gaze data. As additional data sources, he used keystroke logging and participants' self-ratings of the effort they expended on each sentence. In addition, participants were not allowed to backtrack and revise previous sentences nor access online resources. While they had some translation background ranging from beginner to professional level, post-editing experience was not a criterion (Vieira 2016a, 2016b, 2017).

The next larger units of analysis are the translation process phases *orientation*, *drafting* and *revision* based on Jakobsen (2002), among others. While these phases may be performed iteratively, each of them is generally associated with specific cognitive processes: the orientation phase with source text comprehension, the drafting phase with target text production and the revision phase with monitoring and editing. Heidrich (2016, 222–224) also divides the translation process – the target text production

process[23] in her model – into three phases: reception, transfer and production. To these phases, she assigns distinct cognitive activities such as for example knowledge generating, which she relates to the translator's information search activities that are performed during the reception phase (Heidrich 2016, 222). However, such allocations seem to suggest a linearity and sequentiality of phases that do not do justice to their iterative, blending and overlapping nature. While such three-phase-models of the TP definitely have their merits, as for example enabling the distinction between different kinds of revision behaviour (Jakobsen 2002), they seem to have inherited certain cognitivist notions about cognition. Muñoz (2016, 154) therefore questions their usefulness in cognitive translatology. In addition, the models do not seem to have been convincingly supported by observations of professional translators at work. For example, a translator may not read the whole ST before starting with TT production, skip large ST segments when drafting and only insert them during the revision phase, or do information searches merely in the revision phase (for additional examples cf. Asadi and Séguinot 2005). Such variation in behaviour makes operationalisation of the phases difficult, which is necessary for their empirical investigation. This then may distort results for each phase or their comparability.

Hvelplund (2011) used Jakobsen's (2002) model, though, as a starting point for investigating the distribution of cognitive effort during the drafting phase (see also Hvelplund 2017a, 61). He chose that phase as his unit of analysis as he considered it the stage where the translator performs the majority of the cognitive processes that are actually related to translation: Her "overall goal [in the drafting phase] is to create a translation of the SL message in the TL" (Hvelplund 2011, 49). Hvelplund analysed gaze data to calculate whether the participating professional translators and translation students focused more on the ST or on the TT. In order to be able to manage the complexity and effort of data analysis, the participants had to work within predetermined and fixed ST and TT frames and without access to online or other documentary resources. Another example of investigating translator effort during the drafting phase is Hvelplund's (2017a) study on four types of reading: ST reading with or without concurrent TT production and TT reading with or without concurrent TT production. He conceptualises reading as text comprehension that involves the identification, interpretation and organisation of visual sensory information (2017a, 56). He triangulated gaze and keystroke-logging data to identify occurrences of the four reading categories. It seems that Hvelplund equated gazing at the screen as identified by visual fixations as reading while other researchers define reading as a sequence of fixations and saccades. His claim that "reading during translation

23 My translations.

consists of several types of reading with different underlying purposes and activities" (Hvelplund 2017a, 57) seems reasonable. However, caution is called for in ascribing certain cognitive processes such as *ST comprehension* to one type of reading only and *translation* to one distinct phase only. Cognitive processes might well interact and overlap with process activities and the end of an activity does not automatically determine the end of a cognitive process assumed to be related. Therefore, I prefer a wider – but for my research scope still manageable – definition of when *translation* happens: it starts when the translator expresses that she is ready and ends when she expresses that she has finished.

The next larger unit of analysis is the whole translation process. This means that participants perform a complete translation task or again, what the researcher defines as a complete task, which is then analysed with regard to translator effort. A phenomenon often investigated over the whole task is pausing behaviour. Adopted from writing research, *long* pauses have traditionally been associated with higher order cognitive processes such as planning or monitoring and therefore with effortful cognition. However, the discussion about an appropriate pause threshold for translation is ongoing. Studies with different pause definitions have for example been conducted by Dragsted (2012), O'Brien (2006b), Kumpulainen (2015), and Muñoz and Martín (2018). Other process activities can also be investigated over the whole task. Hvelplund (2017b), for example, investigated the three process activities *digital resource consultation*, *drafting* and *end revision*. He used the time that the participants spent on each of the three activities as an indicator of temporal effort (2017c, 72).

He used eye tracking and screen recording to monitor 18 professional technical and literary translators translating two literary and two technical texts. If only the results for the two technical texts are considered, digital resource consultation accounted for on average about 25 %, drafting for 55 % and end revision for 20 % of total task time (2017b, 76). Hvelplund also determined what he referred to as processing effort (as indicated by mean fixation duration) and cognitive load (as indicated by pupil dilation), which are both highest for digital resource consultation. He argues that this result may be explained by the greater variety of tasks associated with information seeking than with drafting and revising (2017b, 77). However, something else might be more relevant: the translator's effort to keep in mind her macrostrategy for the task, to select and evaluate among a vast amount of information presented in non-translation specific resources and to decide on how to integrate it into her TT. Hvelplund's insights that information seeking is an essential part of the translation task and that it involves different types of reading are important. Based on the same data, Hvelplund and Dragsted (2017) reported on the impact of genre familiarity on cognitive effort. The results corroborated their hypothesis

that translating a less familiar text genre is generally more cognitively demanding than translating a familiar one. In other words, the literary translators had significantly longer mean fixation durations than the technical translators when translating a technical text and vice versa. It is possible that the translators had adjusted their behaviour in order to cope with their unfamiliarity with the respective tasks. Hvelplund and Dragsted's study point to an important and ongoing issue in CTS: the choice of source texts and its impact on study results. For my study, I use STs of a text type commonly encountered and on a general topic and make sure that none of the participants is an expert in that specific genre-topic combination.

Krings' (2001) study is also important in the considerations here, as he compared translation effort to post-editing effort with and without access to the ST in the context of complete tasks. As already mentioned, Krings suggested three different types of effort: temporal, technical and cognitive. As a measure of temporal effort, he used processing speed, which he defined as the number of source text words processed per minute (2001, 277). As a measure of technical effort, he counted the number of revisions or edits performed. To investigate cognitive effort, Krings transcribed, coded and analysed the think-aloud protocols and calculated the amount of verbalisation for the different task types. Relative post-editing effort is measured by comparing post-editing speed to translation speed and also by comparing the amount of verbalisation produced during post-editing to that during translation. However, the relation between the effort types is not straightforward, as for example the translation of a short but complex ST may involve high cognitive effort, high temporal effort and low technical effort if the translator does not revise a lot or high technical effort if she does. Lacruz (2017, 387) points out that "[i]n addition to the effect of different translation situations, relationships between the different types of effort will always be affected by individual differences between translators with varying expertise or style". In my view and along the lines of cognitive translatology, Kring's temporal and technical effort can be interpreted as different indicators of the same cognitive construct, i. e. translator effort. Krings' choice of effort indicators made sense in connection with the data collection tools he used. Nevertheless, subsequent studies on effort in post-editing generally have adopted Kring's three types of effort but substituted the use of think-aloud protocols by the use of eye tracking (see e. g. Vieira 2016a or Moorkens 2018).

The relation between translation process activities and underlying cognitive processes. Hvelplund, an expert in using eye-tracking technology in CTS, has expressed caution towards the generally assumed synchronicity between visual and cognitive focus during task performance. He mentions the possibility of unintentional mind drifting (Hvelplund 2014, 209) or, in other words, the possibility that the cognitive focus may shift independently

of eye movement. "While the eye tracker can fairly accurately identify where the eyes are looking, it cannot identify the object of thought" (Hvelplund 2014, 210). In addition, instances of multiple processing in translation have been identified (see e. g. Macizo and Bajo 2006; Ruiz, Paredes, Macizo and Bajo 2008; Balling, Hvelplund and Sjørup 2014). They indicate that connectivist assumptions of sequential and clearly distinguishable processes may not be useful for CTS any more (Hvelplund 2017a, 74). ST-related cognitive processes such as comprehension are not restricted to ST reading and TT-related processes are not restricted to TT reading. We should indeed be cautious with relating observable activities directly to specific underlying cognitive processes. Psychological reality suggests that the translator may already start preparing the TT while reading the ST or that she may indeed compare TT to ST while reading or even typing TT. One possible explanation is the necessary formation of a macro strategy that the experienced translator adheres to and adjusts during the task (Risku 2016, 59).

House (2015, 118), among others, generally questions the suitability of eye tracking, keystroke logging and screen-recording technology to make inferences about underlying cognitive processes. As she put it, can "measurements of observable behaviour really explain the nature of cognitive representations of the two languages, throw light on a translator's metalinguistic and linguistic-contrastive knowledge, and illuminate comprehension, transfer and reconstitution processes emerging in translation procedures [...]? Not really!" I agree with House but not for the same reasons. In my view, translation task activities are triggered by underlying cognitive processes and vice versa and in that respect, the two are indivisible. Nevertheless, concurrency and congruency of activities and cognitive processes cannot be inferred from this premise. As a consequence, single subprocesses such as House's *comprehension, transfer and reconstitution* cannot in fact be illuminated because they do not occur separately and it is impossible to establish where one type ends and another one starts. However, CTS has shown that similarities and differences in behaviour compared across tasks or individuals can tell us whether there are also similarities and differences in the underlying cognitive processes and in their effects and causes, i. e. in cognitive effort and cognitive load.

In this section, I have provided an overview of how translator effort has been conceptualised and investigated in Cognitive Translation Studies in general. As most of the studies dealt with translation into the L1, the next section focuses on research that has investigated translator effort with regard to directionality.

3.5 Directionality and translator effort in Cognitive Translation Studies

Having already discussed research into translation processes and products with regard to directionality (see Section 3.2) I now discuss studies that looked into translator effort as an additional aspect. A study that covered all these aspects is Lorenzo (2002b), whose studies are already referred to in in Section 3.2. Lorenzo (2002b) examined translator effort over the whole task and used the number of keystrokes, task time and the number of pauses longer than three seconds as indicators. She found that those L2 translators who produced the most acceptable target texts also had the most labour-intensive processes (Lorenzo 2002b, 103). Lorenzo (2002a) also reports on the same L2 translators' revision competence during the revision phase. Whereas there is no correlation between *revision effort* – operationalised as the number of revisions and revision duration – and product quality, there is a positive correlation between product quality and the translators' approach to revision: those who only checked for completeness and kept very close to the ST scored lower than those who also revised for TT coherence and text type conventions as well as target audience needs (Lorenzo 2002a, 142). In other words, it was not the amount but the quality of revision that affected product quality. Lorenzo (2002b) noticed that what is generally called the drafting phase varies considerably between translators. While some tend to produce only a scaffolding that contains many untranslated segments, others produce a more or less complete first draft. Lorenzo's description of such individual patterns in professional L2 translation is in accordance with findings on L1 translation patterns as e. g. done by Asadi and Séguinot (2005). Individual differences then of course also have an impact on the types and number of activities performed in the revision phase and thus on its duration. Considering only revisions in the revision phase as a basis for correlation with product quality may therefore be too narrow an approach.

Two studies that also investigated revision behaviour and related it to cognitive effort were conducted by Ferreira (2012 and 2014). In her 2012 study, Ferreira had ten professional translators, five with English as their L1 and five with Brazilian Portuguese as their L1. All of them had experience as bidirectional translators (i. e. from L1 into L2 and vice versa). They had to translate one English and one Portuguese ST of about 250 words on related topics and of the same text type. The texts were introductions to academic papers and had to be translated for publication in an international journal. Five translators performed the L2 translation first and five the L1 translation. Ferreira used keystroke logging and retrospection to elicit data. Her dependent variables and indicators of cognitive effort were *segmentation*, *recursiveness* and *time* spent on each process phase as well

as on the whole task. She defined processing segments as units that were delimited by keystroke-logging pauses of five seconds or longer (Ferreira 2012, 80). For each segment, she then counted the number of recursive movements performed, operationalised as *elimination keys*, *navigation keys* and *mouse actions* recorded in the keystroke log. Ferreira reports that, in relation to the total number of events logged, the translators performed more recursive movements into L2 than into L1. In addition, into L2 they tended to take longer to finish the task and produce more segments, which means they worked on smaller text units. These findings corroborate Ferreira's hypothesis that L2 translation is cognitively more demanding than L1 translation. In her view, the translators had shown themselves able to adapt to task difficulty through *cognitive adaptation* (Ferreira 2012, 87), a concept previously described and investigated by Buchweitz and Alves (2006). In the second study mentioned here, Ferreira (2014) again analysed patterns of recursive movements in TT production combined with translators' comments. Since she assumed that L2 translation and L1 translation differ with regard to cognitive effort, she expected a different number of comment types depending on translation direction. Her eight participants were all professional translators with Brazilian Portuguese as their L1 and English as their L2 and with at least six years of experience in bidirectional translation. In the first session, they translated two texts of about 250 words on similar topics, one into L1 and one into L2. In the second session, they translated two texts of about 200 words on different topics, again one into L1 and one into L2. After every task, they were shown a replay of the recording and "asked to comment on possible difficulties as well as solutions to the problems they encountered" (2014, 116). Ferreira then categorised and counted the comments based on Pavlović's (2010) classification scheme. Ferreira (2014, 125) reported on differences with regard to the two comment types that were most frequently mentioned and concluded that, in the absence of a potential facilitating effect, L2 translation tended to involve more effort than L1 translation.

As another measure of effort, Ferreira (2014) assessed the amount of recursiveness by counting mouse, cursor and backspace movements over the whole process. Results showed that when translating texts on the same topic, the translators tended to perform a lower number of recursive movements in their second task, irrespective of translation direction. This may have been due to a facilitating order effect. However, when translating texts on different topics, the same translators tended to perform a higher total number of recursive movements during L2 translation compared to L1 translation. Interestingly, a strong correlation emerged between the translator's total number of recursive movements and the number of movements that actually resulted in changes to the texts. Individual editing behaviour is thus apparently quite robust (Ferreira 2014, 118). Ferreira

used data from two sources to establish indicators of cognitive effort and combined the results to illuminate the relation between cognitive effort and directionality. However, the study's validity and claims would have been strengthened if measures of relative instead of absolute frequencies had been provided. For example, Ferreira concluded from a higher total number of comments on *lexical problems* in L2 translation than in L1 translation that the translators actually had more lexical problems in the former direction. However, as the numbers are not set in relation to e. g. the total numbers of translation problems, the result's explanational power remains unclear.[24]

Another study that triangulated data from different sources to elucidate the relation between cognitive effort and directionality was conducted by Fonseca (2015). She focused on eight professional translators by having them perform two L2 and two L1 translation tasks with the English-Brazilian Portuguese language pair. Fonseca monitored their TPs with keystroke logging, eye tracking, questionnaires, and free and guided retrospection. She triangulated some methods to identify macro translation units, which were defined as sequences of continuous TT production (based on Alves and Vale 2009, separated by pauses shorter than five seconds). Her assumption was that a translation unit produced in one go without being altered later on involves the least cognitive effort and one that is revised both during the drafting and the revision phase involves the most cognitive effort. She defined four types of translation units and counted them to determine whether the revision of a translation unit and the timing of that activity are related to directionality. Fonseca found that in both translation directions, the most frequent translation unit was the one that was not revised (Fonseca 2015, 120). With regard to the number of translation units involving the highest cognitive effort, there was no significant difference between L2 and L1 translation, and the distribution of translation unit types was very similar in L2 and L1 translation tasks (Fonseca 2015, 120). She concluded that L2 translation may not involve considerably more effort than L1 translation does (Fonseca 2015, 122). However, as the translator profile based on revision behaviour changed for some of her participants when they translated into their L2, Fonseca (2015, 123) took this as an indication that translators may proceed differently based on translation direction. The question remains, though, whether this is due to the fact that they usually translated into their L1 (Fonseca 2015, 119).

24 Ferreira, Schwieter, Gottardo and Jones (2016) measured and compared the cognitive effort that L2 translators spend on the ST, the TT and the internet during the TPs with the use of eye tracking. Their study is an important contribution to research on directionality as it includes the translators' information search activities. However, methodological shortcomings make it difficult to appraise its results.

While the previous studies analysed and interpreted indicators of cognitive effort derived from several data sources, other studies have used only one source. To my knowledge, Pavlović and Jensen (2009) were the first to investigate potential effects of directionality with the use of eye tracking. They related the distribution of visual attention to either ST or TT to the cognitive effort invested in processing either of them. Their indicators comprised *total gaze time*, *average fixation duration* and *pupil dilation* plus *total task time*. They focused on differences between translators with presumedly different levels of translation competence. Therefore, they reported on the results of four professional translators and four final year students of translation who translated one text from L1 Danish to L2 English and another from English to Danish without access to information material. The text stimuli (of around 250 words each) had been made comparable by controlling for factors such as genre, readability, grade level, syllable count and sentence length. The results showed that in both translation directions, the cognitive effort of processing the TTs was higher than of processing the STs (Pavlović and Jensen 2009, 100). However, Pavlović and Jensen's hypothesis that L2 translation involves more cognitive effort than L1 translation was only partially supported. So was their hypothesis that students invest higher cognitive effort than professionals, irrespective of translation direction (Pavlović and Jensen 2009, 106). While it is certainly an advantage to have several indicators instead of only one, it is uncertain whether they actually measure cognitive effort as the findings did not correlate. This is supported by a study by Hvelplund (2011, 21–22). As the results on directionality are not reported separately by professionals and students, they do not inform on professional L2 translators' behaviour. In addition, to link visual focus on the ST *exclusively* to ST processing (reading, comprehension) and focus on the TT only to TT processing (production, revision) in hindsight seems too narrow an approach (Pavlović and Jensen 2009, 95). If we consider that a translator switches frequently times between ST and TT during the task, the notion that she processes more than *meets the eye* becomes feasible. Pavlović and Jensen's (2009, 107) interpretation that processing a ST in one's L1 can be just as demanding as processing a ST in one's L2 may serve as a case in point.

Using Pavlović and Jensen's (2009) methodology as a starting point, V. Chang (2011) investigated cognitive effort over the whole task instead of divided between ST and TT. His participants were 15 Mandarin L1 speakers who had completed a one-year postgraduate translation or interpreting programme in the UK. The participants were asked to translate two texts of 50 words each, one into English and one into Mandarin as quickly as possible without performing any revision and without access to information resources. As V. Chang wanted to minimise the effect of ST complexity on the results, the STs were adjusted to have the following features: suitable

for fourth-graders, easy to read, easy to comprehend, similar in number of words and sentence length (determined by two Mandarin teachers by consensus), and easy to translate. V. Chang added *translatability* of the ST as an additional control for task comparability. As tests for e. g. readability are generally not suited for comparison across languages, he had bilinguals with Mandarin as their L1 rate the two STs with regard to comprehensibility and readability. To determine translatability, he had two Mandarin language or translation teachers with English as their L2 rate the STs on a five-point scale from *very easy* to *very difficult* to translate (V. Chang 2011, 164). Into their L1, the participants produced TTs in Simplified Mandarin or Mandarin Chinese, depending on their geographical origins. Two external evaluators then graded the target texts based on a standard scoring scheme. The TTs were all considered to be of very high accuracy (V. Chang 2011, 168). Rather unconventionally, he reported these scores as behavioural results that served to validate his physiological measures (V. Chang 2011, 168). He measured pupil size, fixation count, fixation duration and task time, which all indicated significantly heavier cognitive load in L2 translation than in L1 translation. Measures of blink frequency, however, did not substantiate this. V. Chang seemed surprised that the L2 target texts scored slightly lower than the L1 target texts as he had done everything to ensure comparability of tasks. However, not letting the participants perform any revisions or information searches may have constrained some of them. While V. Chang (2011, 164) acknowledged that the low and even number of TT evaluators was due to practical constraints, it remains unclear whether they were the same evaluators who had also assessed the STs' translatability. If this was the case, preformed expectations with regard to TT content and variant of Mandarin might have had an impact on their grading. At any rate, it is hard to understand why V. Chang seemed to implicitly assume that ST comparability could be attained and would lead to TTs of the same, objectively assessable quality.

Another study on effort in translation between a language with a logographic script and one with an alphabetic script was conducted by da Silva et al. (2017). They compared the cognitive effort in translation and post-editing tasks in the Chinese-Portuguese language pair. While their 18 participants had had ample professional experience as bidirectional translators, they were only introduced to post-editing just before the research task. Da Silva et al. (2017) used eye tracking and keystroke logging to collect data. They measured four indicators related to the ST area and the TT area on the translators' screens and found that only those related to the latter pointed to a higher cognitive effort for post-editing and translating into the L2 as compared to into the L1.

The last study relevant mentioning is an ongoing study that involves professional bidirectional translators with Polish as their L1 and English

as their L2 (Whyatt, Kościuczuk and Turski 2017). They performed two translation tasks in each translation direction, which were monitored using keystroke logging, eye tracking and screen recording. Effort indicators from process data were then combined with results on product quality. The TTs' quality was assessed indirectly by measuring the amount of cognitive effort that professional proofreaders invest in correcting them (also see Section 3.2). As the proofreaders who corrected the English TTs were not the same as those who corrected the Polish TTs, it would be interesting to see how the effort they invested can be compared. Preliminary results based on measures of text production speed and numbers of long pauses in the translation processes indicate that there does not seem to be a difference in cognitive effort with respect to directionality (Whyatt, Kościuczuk and Turski 2017). The fact that the three-year study is nationally funded reflects the importance of L2 translation into English for and its ubiquitous practice in communities with languages of limited diffusion. In contrast, my study involves two major languages in Europe, which are offered as working languages by many translation training institutions. I therefore have the opportunity to recruit both L1 translators and L2 translators into English and have their TTs evaluated by the same evaluators, which facilitates comparability of results.

3.6 Directionality in Interpreting Studies

The following section is a short digression to directionality in interpreting research, as for professional interpreters, working in both directions is also a common practice (Schmitt et al. 2016; IAPTI's Ethics Committee 2015, 13). Models of cognitive load in interpreting (Gile 2009, Seeber 2011) do not mention directionality as a type of load to consider. In theoretical considerations, the view on whether interpreting into one's L2 needs additional effort on behalf of the interpreter seems to be divided. Those scholars who think that input comprehension involves more effort than output production, prefer to have interpreters working from L1 into L2, and vice versa (Gile 2005). Studies on directionality deal with topics such as professional interpreters' working habits (Schmitt et al. 2016; IAPTI's Ethics Committee 2015), their attitudes and preferences (Audeoud and Haug 2013; Haug and Audeoud 2013), combinations of those (Opdenhoff 2011), and their performance (C. Chang and Schallert 2007; J. Wang and Napier 2015; Opdenhoff 2013). Often, questionnaires are used to collect data. Other studies focus on interpreters' speed of lexical retrieval in word translation tasks (Christoffels, De Groot and Kroll 2006; Chmiel 2016). Interestingly, Chmiel (2016) found that interpreters who only worked into their L2 were not faster in lexical retrieval in their L1 compared to their

L2 but interpreters who regularly worked into both directions were. Based on retrospective interviews performed after interpreting tasks, C. Chang and Schallert (2007) learnt that directionality was not the only aspect influencing performance. As important were metacognitive awareness of problems, audience expectation and other perceived expectations as well as language-specific features. This insight supports Gile's (2005) plea that not only language mastery is relevant but also to consider other aspects such as familiarity with topic, experience and motivation. In summary, it seems that research on directionality in interpreting has frequently involved professional interpreters and their personal opinions. Working into both directions has been treated as a practice that is necessary, manageable and increasingly accepted without much ado. After consulting empirical studies, C. Chang and Schallert (2007, 138) concluded that "[a] review of this still limited pool of available research seems to support both sides of the directionality debate".

3.7 Conclusions from the review of the literature

Overall, research into directionality suggests that translation students tend to expend more cognitive effort than professionals, irrespective of translation direction (Pavlović and Jensen 2011), and more effort in L2 translation than in L1 translation (V. Chang 2011). With regard to professional translators, some results indicate a generally higher effort for L2 translation (Ferreira 2012; Ferreira 2014), some a generally similar effort (Fonseca 2015; Whyatt, Kościuczuk and Turski 2017) and others show similarities and differences depending on the indicator (da Silva et al. 2017). In conclusion, the relation between professionals' cognitive effort and directionality remains unclear. The results for translation product quality and directionality suggest that trainee translators produce TTs of higher quality into L1 than into L2, both when working in teams (Pavlović 2007a) and when working on their own (V. Chang 2011). Professional translators' TTs are more acceptable when they translate into their L1 than when they translate into their L2 (PACTE 2009). With regard to experienced bidirectional translators, the relation between TT quality and translation direction has – to the best of my knowledge – not been investigated.

Especially in a situation where the translator performs a reasonably long task in a laboratory, it is easy to assume that the activities performed are related to the cognitive task of translation. Depending on how the translation process is defined, even taking a break can be seen as being part of the process. In a workplace situation, where long tasks are more likely, the activities performed in order to cope with fatigue could also be seen as an integral part of the translation process. As a consequence, I see different

indicators of effort as different representations of the same cognitive effort and not as different types of effort. I therefore do not distinguish between what others, based on Krings (2001), have called technical, temporal or cognitive effort but instead call it translator effort. Based on the definition by Piolat, Roussey, Olive and Amada (2004, 22), translator effort is the amount of resources a translator requires or invests to perform a translation task. Along the same lines, Vieira (2016a, 9) defines cognitive effort in post-editing as "the amount of mental resources individuals allocate to a task" and Sjørup (2013, 10) defines cognitive effort in translation as "the energy the human mind spends on processing and producing language". These definitions do not distinguish between types of processes or levels of processing but do acknowledge that different types and all levels are included in task performance and that distinguishing between them is taxing, to say the least. If a translator expends a higher amount of effort in one task than in another, this does not necessarily mean that she performed more higher order processes in the former than the latter. Therefore, in my study I concentrate on total amount of effort exerted over the whole task without speculating on the relative proportions of different levels of processing.

Searching for information is considered an integral part of professional translation. Nowadays, there is a multitude of online resources, which are easy and fast to access. They seem to have become the resource type of choice, at least if the translators do not work with translation memory databases. Therefore, the participants in my study will have access to the internet. When investigating information search behaviour, consulting information material has usually been described as accessing external resources while thinking, which corresponds to pausing with observable activities, has often been described as accessing internal resources, i. e. one's mind. However, this dichotomy is not justifiable under the paradigm of situated and embodied cognition and therefore not applicable when doing research in cognitive translatology. Human cognition is not merely enhanced or supported by consulting so-called external resources but is only possible, is only generated in exchange with a human's environment and its artefacts (Muñoz 2016, 155). As a consequence, I suggest that the target text is not ready-made in the translator's mind and then just needs to be typed and any gaps filled with solutions found in other resources. Instead, the translator produces it dynamically and only while in constant and mainly not deliberate interaction with her environment during task performance. As the cognitive act of translating and its manifestation in observable activities during task performance are so closely interwoven, it does not make sense to distinguish between different types of effort. As a consequence, I use translator effort as a synonym for cognitive effort in translation. To investigate translator effort, I use a number of indicators, as has become good practice in CTS. Generally, indicators of translator effort

from all data sources except eyetracking have been treated as indirect ones. However, as there seems to exist a certain ambiguity with regard to which eye-tracking measures are actually directly related to cognitive effort[25], I also use indicators from eye-tracking measures as indirect ones only.

I am interested in determining translator effort at the task level. Although the amount of effort a translator expends during the task is primarily individual, some effects may be found within and across distinctive groups. Therefore, I will control my study with respect to STs and translation briefs: all participants will be presented with the same ST and the same brief in accordance with their translation direction. In addition, I will also control for the task setting and situation by having the participants translate in the ZHAW usability lab. If I then adopt the broad and generous perspective that the participants are comparable with respect to what they bring to the task, translation direction will be the only discerning factor that may cause a difference in translator effort.

As I want to investigate the behaviour of professional translators, I will only include professional translators in my sample. The professional L2 translators in my study represent L2 translation practice into English: they are bidirectional translators. I use methods that have been used and are generally accepted in CTS but do not interfere with the participants usual procedure.

Studies with language students or translation students have often had many participants, perhaps because the researcher-lecturers could draw on their own cohorts (see e.g. Krings 1986). In contrast, studies with professional translators have frequently had to make do with small samples, since recruiting professionals is much more demanding. Nevertheless, quantitative studies with professionals have provided valuable results and researchers have provided assistance to enhancing methodological rigour in the field. Rather than content themselves with samples too small for mainstream analytical methods, they have suggested feasible metrics (Balling and Hvelplund 2015, Mellinger and Hanson 2016). Moreover, in order to ensure the validity of results, standardised participant profiling as e.g. suggested by Muñoz (2012b) can be applied. In my study, I match the participants of each group as closely as possible with regard to experience as professional translators, translation training, and age. An additional means is the use of appropriate measures of statistical analysis. I use Mellinger and Hanson's (2016) book on quantitative research methods in translation and interpreting studies as a reference, among others. However, as it is not

25 Hvelplund (2014) presents an overview on different eye tracking measures. However, the definitions of the concepts the measures relate to, as e.g. processing depth, processing effort or cognitive effort, seem to overlap and remain fuzzy.

possible to control for all confounding variables in research, maybe even more so in research with real people, a balance between practicability and reliability has to be found.

Using cognitive translatology as my research framework has consequences for my study that I would like to explicitly address. Since cognitive translatology ultimately demands studying human translation "when and where it happens" (Ehrensberger-Dow, Hunziker Heeb, Jud and Angelone 2017, 117), balancing requirements of research rigour and ecological validity is an important aspect of my study. Details on how I achieve this are found in the *Methodology* chapter. Another consequence concerns the relation between observable activities and underlying mental processes. I do not differentiate between an external and an internal translation process as two distinct or separate processes (for a different view see Schubert 2009, 19). I rather understand the observable translation process activities as the manifestation of and inspiration for the cognitive process of translation. In addition, I assume that distinct observable activities such as writing, reading, information searching and pausing do not strictly coincide with distinct single underlying cognitive activities such as text comprehension or problem-solving since the latter are neither separable nor discrete cognitive processes (Risku 1998, 207). Cognitive translatology conceptualises cognitive processing as continuous in a parallel and distributed form rather than sequential and discrete.[26]

In the next chapter, I explain the methodological approach that I use to answer the following research questions, which I introduced in Chapter 2.

1. In what ways are the processes of professional L2 translators similar to or different from the processes of professional L1 translators?
2. In what ways are the products of professional L2 translators similar to or different from the products of professional L1 translators?
3. In what ways is the effort of professional L2 translators similar to or different from the effort of professional L1 translators?

26 For an overview of the different paradigms in cognition and their implications for Translation Studies see e. g. Krüger (2015, 275–292) or Risku (2016, 69–100).

4 Methodology

In order to be able to address and answer the research questions, I decided to conduct two related studies (i. e. a main study and a substudy). The main study explores the nature of L2 translation processes and potential effects of directionality on the key process activities, whereas the subsequent, dependent substudy explores characteristics of the final products (i. e. the target texts) of those translation processes and the potential effects of directionality on product quality. A selection of descriptors of both the translation process and the translation product are then combined (i. e. triangulated) in order to gain insight into the effort that the translators invested to perform the task. This triangulation of data necessitates a mixed methods approach, which overcomes the traditional divide between qualitative and quantitative research paradigms and can also be described as methodological triangulation (Robson 2002, 174–175). Another advantage of this approach is that the different data sources can be used to enhance interpretability (Robson 2002, 371).

In this chapter, I present the overall study design (Section 4.1) and the designs for the main study *Directionality and the translation process* (Section 4.2) and the substudy *Directionality and the translation product* (Section 4.3). I describe the participants, stimuli and tasks as well as the data collection methods and procedures. I also explain the operationalisation of the concepts under investigation and discuss central ethical and methodological issues. Results from the main study and the substudy are triangulated to build indicators of translation effort, which are introduced in Section 4.4. In the last Section 4.5 of this chapter, additional variables are presented, which will be used to investigate relations between characteristics of professional translation.

4.1 Overall study design: Mixed methods approach

Translation process research (TPR) has gained momentum with the introduction of technical instruments to capture the complex process with means additional to human observation. This in turn has been facilitated by the translators' use of computers to do their work. Multiple sources of data allow investigations of the translation process from different perspectives and (partially) compensate for limitations of the individual instruments. This improves the robustness of a study (see e. g. Keyton 2014). As can be seen from the studies discussed in the literature review, TPR has had a tradition of taking advantage of combining multiple methods (Ehrensberger-Dow 2018) and triangulation became well established early

on as a methodologically valid option to study translation processes (see e. g. the volume on the topic edited by Alves in 2003). Both my main study and the substudy apply a concurrent convergent approach (Creswell and Plano Clark 2011): qualitative data and quantitative data are collected at the same time and then, in the analysis, combined to form reliable indicators of translator effort and inform the relations between product and process with regard to translation direction. An overview of the overall study design is presented in Figure 1. The different types of data elicited in the main study are listed in Table 2 in the Section 4.2.4 *Main study data collection methods*. The various types of data elicited in the substudy are presented in Table 4 in the Section 4.3.4 *Substudy data collection methods*.

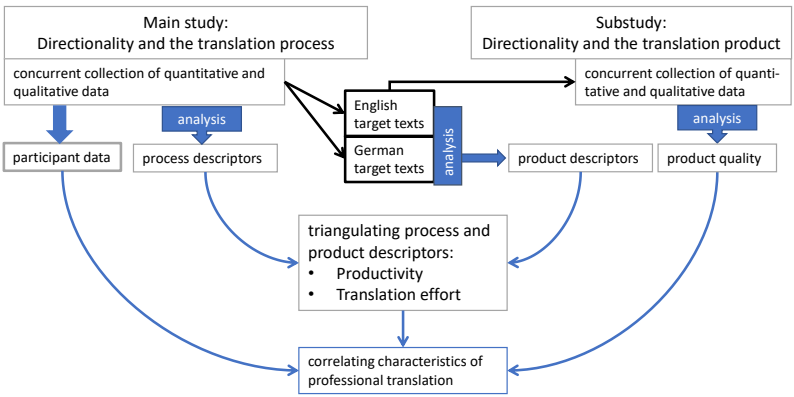


Figure 1: Overall study design: mixed methods approach

I enhance the validity of results by analysing one source of data (e. g. keystroke logging) with respect to results from another source (e. g. screen recording) to complement 'gaps' encountered in the interpretation of results (Kuckartz 2014, 58). An example is to examine the eye movements – represented as gaze patterns in the screen recordings – during long pauses in keyboard activity. Another example is the translators' qualitative accounts of how they usually proceed (i. e. interview data), which can complement and/or corroborate findings on their process behaviour.

Creswell and Plano Clark (2011, 277) as well as Kuckartz (2014, 57) differentiate between mixed methods and multimethod research. They claim that in a multimethod approach, the researcher uses multiple methods to collect either qualitative or quantitative data, whereas in a mixed methods approach, both types of data are collected. Alternatively, Halverson (2017) uses the term multimethod as superordinate for the combination of methods, irrespective of data types. What both designs have in common, though, is their use of triangulation. Triangulation can be performed at

the different levels such as for example at the level of method, data, results, theory or investigator (Flick 2011, 14–16).

In this study, I collect data sequentially in two rounds. During the first round, the main study, data on the translation processes are collected together with participant data. Methods include interview, recording of computer screen activities, verbal reports, logging of keyboard and computer mouse activities and text files (see Table 2). For the second round of data collection, the substudy, the translation product produced during the first round is used as stimulus for quality assessment. Methods include questionnaire, ranking, word count in Microsoft Word and readability determination (see Table 4). The quantitative and qualitative data collected is analysed accordingly, although some results of the latter are quantified to allow certain comparisons to be made. Then I triangulate some of the results to provide more insights into certain concepts than would have been possible with a single indicator and in the last part, I draw relations between them. The results of the main study are reported in Chapter 5 *The translators' processes*, of the substudy in Chapter 6 *The translators' products*, of the triangulations in Chapter 7 *The translators' effort* and of the relations in Chapter 8 *Relations between characteristics of professional translation*. The next section presents the methodology of the main study in detail.

4.2 Main study: Directionality and the translation process

The main study investigates potential impacts of directionality on the translation process. To do so, three groups of professional translators that differ with respect to directionality are included and their processes recorded as they translate into their L2 and/or L1, as explained below.

4.2.1 Main study design

With German and English being the languages of the translation versions under investigation, the following combinations for participants would theoretically be possible: unidirectional translators who translate only from German into their L1 or L2 English or those who translate only from English into their L1 or L2 German; German-English translators who regularly work into both their L1 English and L2 German (i. e. English bidirectional); and English-German translators who regularly work into both their L1 German or L2 English (i. e. German bidirectional). As ecological validity is a central prerequisite of this study, only professional translators who actually use German and English as their working languages were considered as

participants. In recruiting, it proved to be impossible to find English bidirectional translators who work into their L2 German in the catchment area I had access to.[27] From a demand point of view, this makes absolute sense in German-speaking countries, which have an adequate supply of English-German translators working into their L1.

This reality meant that three groups participated in the main study: German bidirectional translators (i. e. into L1 and L2), English unidirectional translators, and German unidirectional translators (see Table 1).

Group	Translation version	Translation direction	Participants	Group size	Translation task	Source text
German bidirectional translators	German-English	into L2	BiDir 1-6	n=6	BiDir_L2	Wale
	English-German	into L1			BiDir_L1	Whales
English unidirectional translators	German-English	into L1	UniEnglish 1-6	n=6	UniEnglish_L1	Wale
German unidirectional translators	English-German	into L1	UniGerman 1-6	n=6	UniGerman_L1	Whales

Table 1: Quasi-experimental design of main study: Participant groups and translation tasks

The first group, the German bidirectional translators (Bidir), have German as their L1 and are used to working in both directions (i. e. German-English and English-German). The second group, English unidirectional translators (UniEnglish) have English as their L1 and translate exclusively into their L1 (i. e. from German into English). The third group, German unidirectional translators (UniGerman), have German as their L1 and also translate exclusively into their L1 (i. e. from English into German). All three groups are the same size (n=6). Whereas the Bidir group performed two translation tasks – one into their L2 (Bidir_L2) and one into their L1 (Bidir_L1) – the UniEnglish group performed only one task (UniEnglish_L1) as did the UniGerman group (UniGerman_L1). This design allows for intragroup as well as intergroup comparisons (see Figure 2). Intragroup comparisons are performed between members of the Bidir group – that is they concern the same translators but different source texts, different language versions, and different translation directions – in order to investigate whether patterns are related to directionality or to individual profiles or preferences. The design also allows for intergroup comparisons between the Bidir and

27 The member database of BDÜ listed some English L1 translators who offered English-German translation. However, many gave an address in the UK or the U. S.

the UniEnglish (same source text, same language version but different translation direction), on the one hand, and between the Bidir and the UniGerman (same source text, same language version and same translation direction) on the other.

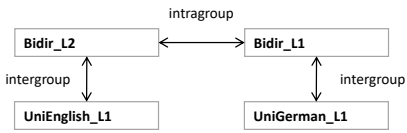


Figure 2: Group comparisons in main study

The study design is quasi-experimental, as it does not control for all variables as would be necessary in a strictly experimental design and does not randomly assign participants to specific experimental groups (for a disambiguation between experimental and quasi-experimental design see e. g. Mellinger and Hanson 2016, 7–8). This type of purposeful group allocation is realistic for many studies on human behaviour (Balling and Hvelplund 2015, 170) and common in translation and interpreting research (Mellinger and Hanson 2016, 7–8). The task and the environment are created exclusively for the purpose of this study with the goal of simulating a naturalistic translation environment in order to be as ecologically valid as possible.

The combination of methods and tools I use is well established and tested in CTS: the triangulation of data and results follows good practice as it allows for the complexity of the translation task and for the opportunity to investigate phenomena from different perspectives. Hansen (2006) corroborated its usefulness in her study on sources of error that she detected in non-professionals' translation processes. She also triangulated process and product data to report on their consequences. What is new in my study, however, is that my participants are professional translators who work regularly in the language versions and translation directions I ask them to perform in. Moreover, I let them proceed as they would usually do. The next section presents the participants in detail.

4.2.2 Main study participants: Professional translators

All participants are professional translators. I chose a pragmatic definition of professionalism, which has been suggested by Jääskeläinen et al. (2011) and Pym (2011): Translators are considered professionals if they earn at least part of their living from translating. This means that they see themselves as translators and are accepted as such by their employers and/or clients on the translation market. The fact that they have been working as translators

for some time allows inferring that they have produced work with an appropriate level of quality. An additional requirement was that they were either trained as translators or accredited as such by a professional body. The following sections describe the two main participant groups, bidirectional and unidirectional translators.

Bidirectional translators

The first main group of participants (for all details listed per participant see Appendix A) consists of six bidirectional translators (Bidir1-Bidir6), who were recruited in Switzerland and in the south of Germany in the context of this study. In order to qualify, participants needed to have done their primary formal education in Germany, Austria or the German-speaking part of Switzerland. Furthermore, they needed to have started to learn English only at secondary school level and none of their parents had English as their L1. Three of the bidirectional translators have Swiss German and three have Standard German as their L1.[28] One of them was raised bilingually with Croatian as her second L1, whereas all the others were raised monolingually. Four of the six bidirectional translators have a translation degree. The other two have a different university degree. At the time of data collection, five translators were working freelance and one as staff translator. Three of them had additional jobs: one as freelance interpreter, one as administrator and one as manager in a translation agency. Their self-estimated engagement as translators ranged from 40 % to 100 %, and their translation work into L2 ranged from 20 % to 95 % of the total workload. Their experience as professionals extended from 10 months to 13 years. While all translators were used to working with word-processing software, three also used computer-aided translation tools. Four reported that they were definitely touch typists and two said they were more or less so.

Participants were recruited via professional and personal networks. Graduates from the ZHAW Institute of Translation and Interpreting were also contacted. It was challenging to find participants for the group

28 I subsequently use *German* for both Standard German and Swiss German as Swiss German is generally not seen as a language of its own but a variety. In the German-speaking part of Switzerland, Standard German is the main language of instruction in the educational context. Apart from that, it dominates in written communication settings, while Swiss German dominates in oral communication settings. As I investigate the former, I do not differentiate between participants with Swiss German and those with Standard German as their L1. For an investigation into whether being a speaker of Swiss German has an effect on Standard German speech production in simultaneous interpreting see Keller (2018).

of bidirectional translators, perhaps because of the prejudices towards L2 translation within the profession (as explained above) and also because participation involved travelling to Winterthur. Every participant was invited with an e-mail explaining that the focus on the study was on the translation process, that participation was voluntary and that all data would be anonymised.

Unidirectional translators

The second main group of participants were unidirectional translators, divided into two sub-groups depending on target language: six with English as their L1 (UniEnglish1-UniEnglish6) and six with German as their L1 (UniGerman1-UniGerman6). They were selected from the pool of participants recruited within the longitudinal study Capturing Translation Processes (CTP n. d.) to match the Bidir group as closely as possible with regard to experience, training, age, and job situation. They all lived in Switzerland, except one who lived in Singapore at the time of data collection. All of them were raised monolingually and they started to acquire their L2 English or L2 German (i. e. their source languages) at secondary school level at the age of 13 to 15 or 11 to 17, respectively. Of the six unidirectional German-English translators (in the following referred to as UniEnglish_L1), all are staff translators. All work full-time and have a translation degree except one, who has another degree at university level. Their experience as translation professionals ranges from eight months to eleven years. Of the six unidirectional English-German translators (in the following referred to as UniGerman_L1), five are staff translators and one is a freelancer. They had been working as translators between two and eleven years at the time of data collection and all were working full-time, except for one who was working part-time on a 50 % contract. All have a translation degree. All 12 participants of the unidirectional groups work with CAT tools. Nine participants reported that they were definitely touch typists, one more or less and two not at all. For all details listed per participant see the table in Appendix A.

In a previous study, I reported on the self-concepts of professional bidirectional translators compared to unidirectionals' (Hunziker Heeb 2016). The participants were the ones I have just introduced. My analysis in that study was based on their retrospectively reported foci of attention during the translation processes. Both the bidirectional and the unidirectional translators showed proficiency in juggling multiple concerns and responsibilities during task , which points to a well-developed translatorial self-concept. If we adhere to *self-concept* as an important component of translation competence, both groups display a comparable level. The following section informs about stimuli and tasks.

4.2.3 Main study stimuli and tasks: Source texts and translation briefs

The two source texts (STs) used in this study are of the same genre (newspaper reports) and on a similar, general-interest topic (the stranding of whales). They both contain typical stylistic features and various types of potential problems for translation (cf. PACTE 2009, 212–216).

The German source text *Wale* and the English source text *Whales* (see Appendix A) are both short extracts of about 100 words from articles that had appeared in daily newspapers in the respective source culture, and which are still available online as of writing. The German ST was extracted from an article by Stephanie Kusma called "Strandungen von Walen – seit Jahrtausenden rätselhaft" that appeared in the Swiss quality newspaper *Neue Zürcher Zeitung* on 29 April 2009. From the original eleven paragraphs and four subheadings, the second paragraph was shortened to function as the ST for the German-English version. As the title, the first part of the title from the original article was chosen. The German ST consisted of 96 words in total. The English ST is an extract from an article by Mark Townsend called "Whales at risk in sonar sea exercises" that first appeared in the British newspaper *The Observer* on 8 August 2004. From the original twelve paragraphs, the fifth paragraph was chosen as the ST for the E-G translation. Its first sentence was shortened in order to make sense as an introductory sentence of a self-contained short piece of news and to decrease the total number of words to 95. The original title was retained as the title of the ST. It was important to keep the changes to the original texts to a minimum in order to preserve their authenticity.

Although the translation of newspaper articles may not be that common in professional practice, the texts were chosen so as not to favour any of the participants. Intended unfamiliarity with the subject matter meant that the translators needed to search online resources for information, which was easily accessible on the internet. As none of the translators rejected or interrupted the task, it can be assumed that they handled it in a similar way to their usual approach to translation work.

The translation briefs accompanying the source texts simply stated that the TTs were intended for a comparable publication in the target culture (see Appendix A). It was left to the translator to decide whether the envisioned target audience for the English TT was British, US American or generally international with English as Lingua franca and whether the audience for the German TT was Swiss German, German or Austrian.

4.2.4 Main study data collection methods

For my mixed methods approach, I used the following quantitative and qualitative methods for data collection: screen recording (SCR), keystroke logging (KSL), eye tracking (ET), cued retrospection (RVP), semi-structured interview and questionnaire. All the methods used for data collection in my study allow non-invasive observation of translator behaviour. They are well-established techniques in cognitive translation studies. A data overview is presented in Table 2, and each data collection method is explained in more detail below.

Type of information (main study)	**Data collection instrument**	**Form of data**	**Type of data**
Personal background of translators	first semi-structured interview	transcript	qualitative
Typical work pattern	first semi-structured interview	transcript	qualitative
Translation processes recorded at the usability laboratory	screen recording (SCR) keystroke logging (KSL) eyetracking (ET) Microsoft Word	video log log target texts	qualitative quantitative quantitative qualitative
Cued retrospection	screen and audio recording	transcript (RVP)	qualitative
Effects on text processing	second semi-structured interview	transcript	qualitative

Table 2: Overview of data collected in main study

Screen recording (SCR)

Screen recording is a technique using software that is installed either on the user's computer or on a mirroring computer. It records all the actions happening on the screen, providing information on which applications have been used, which internet sites have been accessed, and how the target text has evolved. I used the screen-capturing program Camtasia Studio (TechSmith n. d.) as it does not affect computer performance, irrespective of which other programs the translator uses. The screen recorder was installed directly on the participant's computer in the usability laboratory and operated by the technician.

Keystroke logging (KSL)

Keystroke-logging software allows the recording of all keystrokes and mouse actions together with time stamps without interfering with normal computer use or other applications. For this study, Inputlog (Van Waes and

Leijten 2006) was used and installed directly on the participant's computer in the usability laboratory and operated by the technician, was used. Inputlog was initially developed to track writing processes but has also been used in translation research (e. g. Daems, Carl, Vandepitte, Hartsuiker and Macken 2016). The text can be produced directly within Microsoft Word, which is a text-editing environment that most translators are used to working in. Moreover, the screen layout is not fixed and entries in the browser window are also logged. In the Capturing Translation Processes project from which I extracted data to form my comparison groups, Inputlog version 3 had been used, as it was the only stable version at the time, which was an important requirement for the experiments. For the data collection of the Bidir translators, version 5 was used. By then, the software had been enhanced with report generating facilities, for example on revisions. Unfortunately, the former data format of version 3 was not compatible with the newer version. Therefore, I had to create most analyses and reports manually via generated Excel files and semi-automatically with an open-source visualizer (Ehrensberger 2014).

Eye-tracking (ET)

Eye-tracking software records the translator's eye movements within a predetermined area. For professional translation, the translator's computer screen represents an adequate area of interest that can be investigated in a laboratory setting such as the ZHAW usability lab (Figure 3). For the study presented here, a Tobii T60 eye tracker and Tobii Studio software were used (Tobii 2019). Tobii T60 eye trackers are integrated in a 17-inch flat screen monitor and operate at 60 Hertz, which means that it records eye movement data every 16.5 milliseconds. The translator's conscious and unconscious eye movements are tracked by capturing reflections of infrared light from the eyes and their position located by providing a horizontal and vertical coordinate for each sample for both eyes. This results in a large amount of data from which fixations (when the eyes are relatively stationary) and saccades (when eyes reorientate and change their focus of attention) can be calculated. The use of this type of eye-tracker ensures a non-invasive data collection method as the participant is free to move her/his head, which would not be possible with a head-mounted eye tracker. The settings used in this study are comparable to the ones tested and proposed by the EYE-to-IT project (Jakobsen and Jensen 2008) and by Hvelplund (2011). For more details see Appendix A.

Figure 3: Usability lab setting with participant in the background and observers in the foreground

Cued retrospection (RVP)

Cued retrospection, also known as 'retrospective verbalisation', which results in retrospective verbal protocols (RVP), is a think-aloud technique used right after the translation process. In my study, the participants were shown the screen recording of their process and asked to comment freely. The integrated gaze replay, visualised as dots for fixations and lines for saccades, acted as a cue. As is the case with concurrent verbalisation, retrospective comments cannot be taken at their face value as they do not reflect the actual cognitive processes. In retrospection, thought and decision-making processes are reproduced from memory and therefore reconstructed (cf. Saldanha and O'Brien 2013, 125). The comments can be seen as representing what the commentator thinks worth mentioning while watching the recording. The amount and accuracy of comments depends, for example, on memory capacity or the willingness to report as mentioned by Englund Dimitrova and Tiselius (2014, 177–200). However, the use of retrospective data is "an excellent way to study the way subjects construct what happened in their minds when they were translating. This could shed light on the way translators envision their tasks" (Muñoz 2010, 181). In this study, analyses of the protocols will be used mainly to illustrate process behaviour. In a precursor study, the RVPs were analysed quantitatively to investigate the translators' self-concepts (Hunziker Heeb 2016).

Interviews

Two semi-structured interviews were conducted (see Appendix A). They are based on protocols that have been used in similar research. The first interview, which was conducted between the translation task and retrospection, provides background information on the participant's language biography, training, work experience and approach to translation tasks. The second interview, which was conducted after retrospection, provides information on potential effects of the lab situation and any 'Eureka' moments on their own process procedures. The members of the Bidir group were also asked whether directionality usually affects how they proceed.

4.2.5 Main study data collection procedure

Each participant performed one or two translation tasks at the single computer workplace in the usability laboratory of the ZHAW Institute of Translation and Interpreting. The six bidirectional translators visited the usability laboratory between June 2013 and February 2014, except for Bidir5, who had already performed her first task (English-German) within the Capturing Translation Processes project in 2012 and who performed her second task (German-English) in July 2013. All six translators translated both the German and the English STs; three of them started with the German ST and the other three with the English one. The twelve unidirectional translators visited the usability laboratory between July 2010 and September 2011. They only translated one ST — the German one if they work from German into English (UniEnglish_L1) or the English one if they translate into German (UniGerman_L1).

For a detailed account of introducing the participant to the translation session at the usability laboratory and the setting up of her workplace see Appendix A. The goal was to make the participant feel welcome and minimise potential effects of affective factors such as nervousness or stress on her performance.

The participants were allowed to use online resources for information searches, just as they would normally do at their workplace. There was no time limit set for finishing the task. Afterwards, while the technician extracted the screen recording as a video file, the researcher conducted the first semi-structured interview to gain information on the participant's language history and typical work pattern (see Appendix A). The participant was then asked to sit in front of a laptop and comment on her/his translation process which had been enhanced by the visualised eye-tracking data (i.e. fixations as orange dots and saccades as lines) overlaid on the recording of the screen activities. The recordings were shown to the

participants in full length, and they were asked to comment in the language of their choice. During this time, the researcher was seated next to the participant but beyond her/his field of vision so as not to distract her from the recording and not to be perceived as a dialogue partner. The play-back of the initial warm-up task helped the participant become familiar with seeing the visualised eye-tracking data and the activity of commenting on a recording of their activities. The participants were not asked to comment on anything in particular, but were encouraged to talk freely about what came to their mind when watching the video. If they paused for longer than a minute or two, they were prompted to continue. In order to align their comments to the activities on the screen, the commentaries were recorded as an audio track overlaid on the process video. Whereas the bidirectional translators then proceeded to the second translation task, at this point the unidirectional translators participated in a second semi-structured interview (see Appendix A). It was conducted at the end of each data collection session to capture any effects caused by the setting and insights the participants had as a result of watching their own translation processes. The data was then anonymised and stored on an access-restricted distinct partition of the ZHAW server system. Each participant was remunerated for participation and her travel expenses were paid for. This was done to compensate for their invested working hours and also to make the task comparable to a real translation commission.

4.2.6 Main study operationalisations

In order to be able to analyse the collected data and to address the research questions, the concepts need to be operationalised and the units of analysis need to be defined. In this section, this is done for the main study. The variables needed to analyse the translation processes in detail are presented as measurable units of analysis (cf. Saldanha and O'Brien 2013, 24). The section begins with the translation process itself and then explains each process descriptor in turn.

Translation process

As already mentioned in Chapter 2, I define the translation process (TP) in a narrow sense: It starts when the translator is seated at the computer prepared to start producing a TT and it ends when she has finished producing the TT. Its start is operationalised as the time the translator – following the instruction on screen – hit the space bar on her keyboard to indicate that she was ready to start the task. This time stamp is easily determined from the KSL data. The end of the TP is operationalised as the

time when the technician stops the logging software by pressing a set key combination minus three seconds. The three seconds represent a feasible default time lapse between the translator informing the technician (which had not been logged as there was no audio recording) that she had finished and the technician terminating the experiment. I chose the deduction so as not to penalise short TPs on which the time lapse would have had a greater impact than on long TPs.

In order to describe individual TPs and facilitate comparisons, process descriptors are defined and operationalised in the following section.

Process descriptors

To be able to describe and compare the translation processes, six descriptors were chosen: Process duration, character count, writing, revising, information seeking, and pausing. All descriptors are sourced in the previously defined TP and assumed to be task related. This is the advantage of a lab setting as compared to a workplace study where the translator may decide to interrupt the TP and perform non-task related activities. The first two descriptors *process duration* and *character count* are global process measures. They are used to get a first impression of the TTs and to allow for comparisons. The other four descriptors represent the main activities that are observable on the translator's computer screen during the translation task: *Writing*, *revising*, *information seeking* and *pausing*. The main source of data for categorisation is KSL as it provides exact time stamps as well as content. Additional sources are SCR and RVP. The four process descriptors constitute the translation process on a macro-level. They are explained in detail below.

Process duration

Process duration is defined as the time it took the translator to accomplish the task, i. e. to produce a target text she thought fit for its purpose as stated in the translation brief. Process duration is extracted from the KSL data and represents the duration of the TP as defined above. The time needed to solve any technical problems or when the technician had to provide assistance was deducted from the total task time. The starting and end times of those interruptions were determined from the translators' comments during retrospection and then confirmed by consulting the screen recording.

Character count

Character count is defined as the total number of characters produced during the TP. It comprises characters (i. e. letters, numbers, punctuation marks and spaces) that were produced by hitting the respective keyboard keys either in the TT document or in the online browser. Excluded from the count are the keys needed to produce uppercase letters, i. e. LShift, RShift and CapsLock as their use was logged differently by the two software versions in use. Also not included are character keys pressed in key combinations (e. g. Ctrl+C) to execute textual operations such as copying and pasting or keyboard activities performed to move within the text or between windows (e. g. CursorDown or Alt+Tab). A global count of all keyboard and mouse actions had previously also been considered as a coarse global measure on the TP. However, as two different versions of the keylogger Inputlog were used, it could not be guaranteed that the logs contained comparable information in that respect.

Writing

Writing is defined as a sequence of successively produced characters or single characters that are added to the right of any already existing TT or mark the beginning of TT production. A writing sequence starts with the hitting of the first key on the computer keyboard and ends with the release of the last key and includes any characters and spaces produced in between. It is delimited by one of the other three process activities (see below). Writing is not performed throughout the whole TP but ends when a first draft of the TT is completed. The information about writing is obtained from KSL, and from SCR or RVP if needed.

Revising

Revising is defined as the deletion or insertion of a single or a series of characters in the translator's own TT.[29] A revision sequence starts when the first character is made to disappear or appear and ends when the last character has been made to disappear or appear, respectively. While a deletion can be performed at the very end of an emerging TT, an insertion is by definition enclosed by existing text. Depending on how revising is executed, its duration may be short despite comprising long strings of text (e. g. if a sentence is selected, cut and then pasted into another position). The duration of revisions depends on how the translator handles the

29 I did not include revisions of search terms in this category.

keyboard and the mouse. An obvious example: deleting a whole sentence with the backspace key may take much longer than selecting it with the key combination *Ctrl+LeftMouseButton* and then deleting it by pressing the *backspace* key once. A revision sequence is delimited by any of the four process activities, i. e. also by a revision that is performed at a different position in the TT. The information about revising is obtained from KSL, and from SCR or RVP if needed.

Information seeking

Information seeking is defined as the activity when the translator accesses and uses her internet browser in order to perform information searches . In the ZHAW usability lab, participants can access the internet and additional resources provided by the ZHAW library via intranet login. Nevertheless, in my study, all types of resources are subsumed under the expression *online resources*. The *information seeking* activity starts when the translator opens the browser for the first time or brings the browser window to the foreground on her computer screen. She does this either by clicking on the browser tab in the task bar or using the Alt+Tab key combination. Information seeking does not end with switching to or opening an additional online resource or with modifying a search term, under the premise that the search topics are related. It ends, however, when the translator leaves the browser window, usually by clicking on the ST or TT icon in the task bar or by pressing *Alt+Tab*. Consequently, if a translator performs a search with a search term identical or similar to one that she has already used earlier in the process, it is still counted as an additional search incident.

Some of the information searches are single-step and others multi-step: a single-step information search consists of one search query in one online resource. As soon as the translator modifies the search term and submits a new query within the same resource or enters the same or a related search term in a different resource, this counts as an additional step. Steps follow each other immediately and are related to the same topic.

Based on Gough (2015, 114), I categorise online resources according to their principal nature and not according to their potential use. However, I do not distinguish between electronic and online resources but between termino-lexicographic and corpus-based ones (for definitions see Gough 2015, 115). This results in the following categories used in the present study:

- bilingual dictionary (e. g. leo.org)
- search engine (e. g. google.ch)
- parallel corpus (text corpora in both SL and TL, e. g. linguee.com)
- encyclopaedia in TL (e. g. Wikipedia sites in TL)
- encyclopaedia in SL (e. g. Wikipedia sites in SL)

- monolingual dictionary in TL (e. g. dictionary.com)
- parallel text in TL (e. g. websites with news reports on beached whales in TL)
- spell checker (integrated into the Microsoft Word word processing program)
- synonym finder (integrated into the Microsoft Word word processing program)
- other (e. g. user forum)

In her study, Gough (2015, 147) excludes spell checkers from her categories as she considers them to be tools rather than resources. She argues that spell checkers are applied to the whole TT rather than to specific translation problems. However, this need not be the case. Therefore, I included them in my study as a separate category just as I did with synonym finders.

Pausing

Pausing is defined as a sequence of more than five seconds when there is no activity logged in the KSL. While the translator pauses in this sense, she may still perform activities that leave no trace in KSL such as reading TT, skimming a hit list, looking at the ceiling or taking a sip of water. Every occurrence of one of the other three process descriptors writing, revising and information seeking ends at the latest when another one starts. This is also applied to pausing. As a consequence, pauses that occur during information searches are excluded.

4.2.7 Main study ethical issues: Professional translators as participants

When professional translators are asked to participate in research, some ethical issues have to be addressed. However, while there is a range of articles on good practice in empirical and experimental translation research (see e. g. Neunzig 2011 or Balling and Hvelplund 2015), there still seems to exist a lack of publications on research ethics in the field. While in some European countries such as Ireland, every research project including humans as participants has to undergo an evaluation by an ethics committee, in countries such as Switzerland only the collection of health data or research that includes vulnerable people like patients or children has to be approved by an ethics board. Nevertheless, ethical research concerns or solutions hardly ever find their way into academic publications. For instance, in T&I research papers, it is often not mentioned how the participants were recruited and what standards of research ethics were

applied. O'Brien (2010, 259–260) considers the ethics approval procedure to be an important step in preparing a study as it "requires the researcher to state in a comprehensible way the purpose of the research and precisely what is required from participants [...and it] also sets out participants' rights" (2010, 259). Hekkanen (2007), then a PhD student, focuses on ethical issues when doing research on literary translation. She argues that in her field anonymisation does not really work as within the small circle of literary translators it would be easy to identify who is meant. Therefore, it would "be beneficial for TS researchers, particularly young researchers, to have translation-relevant ethical issues separately discussed, collected and made available for the TS research community" (Hekkanen 2007, 234).

Studies in CTS with professional translators as participants are still scarce as they involve a large amount of planning and time commitment. Good practice in research implies safeguarding the professional translators' interests as well as preventing and eliminating any reputational risks for their clients and employers. This includes selecting tasks that are in line with the translators' work ethics and professional pride.

4.2.8 Main study methodological issues

As with any study that involves people, confidentiality and non-recognisability have to be guaranteed. Other compulsory requirements such as participants' informed consent, right to withdraw, safe storage of data, anonymisation of data and results, and access to study results are met in this study. Another challenge regards the handling of a large amount of data, which entails finding a balance between requirements of research rigour, ecological validity and pragmatic considerations such as the time spent at the laboratory and the time needed to analyse the data. A case in point is the sampling technique I used. Participants for the main study were recruited by so-called purposive sampling through professional and private networks and within two larger studies (see e. g. Saldanha and O'Brien 2013, 34). This technique was deemed feasible as the methodological constraints required that the participants be professional translators with English and German in their language portfolio and within reach of the ZHAW usability laboratory.

Another example is the balance between the experimental situation and ecological validity. Muñoz (2010, 181) stipulates that cognitive translatology favour "naturalistic data collection activities" so as not to produce non-sensical results. To comply with this, a standard computer workplace was set up in the ZHAW usability laboratory. Translators had access to the internet with all the online resources as well as to the resources offered by the ZHAW library within the ZHAW network. As all participants stated that they use online dictionaries on a regular and frequent basis, the lab set-

up allowed them to access the same resources or comparable ones to those that they normally access.

As I mentioned in the introduction to the paradigm of situated and embodied cognition (Chapter 3), translation processes are very individual. Various external and internal factors such as the job situation or motivation can influence the task at hand. Those factors cannot be controlled for in an experiment. The factors I controlled for were: the setting in the usability laboratory, the script for the translator's visit to control the sequence of activities, a warm-up task to familiarise the participants with the equipment and environment, the same feasible, authentic source texts, and the same translation brief.

Efforts to standardise studies in CTS have been initiated (cf. Muñoz 2012b) and profiling participants is recommended as one of the measures. While I performed some profiling (see Section Main study participants 4.2.2), I decided against testing the language and typing skills of my participants as I considered it would be an infringement on their integrity as professionals. Moreover, it would have prolonged their visit to the usability laboratory considerably. Instead, I decided to rely on their self-declaration regarding typing skills. As for their language skills, I presumed them to be at an adequate level in their working languages, as otherwise, they would not be able to earn money as translators. Moreover, the validity of language tests with regard to translation skills is questionable. Instead, profiling was done via an interview on language acquisition and education. With this, the risk of exposing individual translators as being incompetent or inefficient is avoided, as is supported by O'Brien (2010, 260). My study on the self-concept (Hunziker Heeb 2016) of the same 18 translators showed that they exhibited professional behaviour in the sense that they attended to various types of concerns on the continuum from word level to intended target audience. As a consequence, I weighted the professional behaviour higher than mere language skills.

The laboratory setting for the data collection was a compromise with respect to ecological validity in the interest of comparability. Another compromise was not to allow the use of CAT tools for the translation tasks. At their workplaces, most of the participants use CAT tools, which provide them with suggestions for items that have already been translated within another translation job. They also have the possibility to search for solutions in a concordance tool. These sources probably represent their prime points of departure for information searches. However, all of the participants were also used to translating without CAT tools. Some of their tasks involve short translations or translations that do not fit the client's typical portfolio of text types or topics. For information searches, they then access the internet. A study with the same pool of participants as the unidirectional staff translators in my project compared the translation process activities in the

usability laboratory to those of the same translators at their workplace. It revealed very similar percentages of the four core activities writing, revising, information seeking and pausing (Ehrensberger-Dow 2014, 377–378). In the workplace processes, the automated insertion of accepted translation matches offered by the CAT tool were categorised as writing activities. That study's results substantiate the assumption that translation processes in a controlled setting can be comparable to those done at the workplace.

Another aspect that needs consideration when doing translation process research in a lab setting, is the length of the source texts. The two STs used in this study were 96 words and 95 words long, respectively. This was considered suitable because the professionals had to be able to do the translations within a reasonable period of time in the usability laboratory. In addition, each translation task was followed directly by a cued retrospection and an interview. In order to keep the translators' stay at the usability laboratory to an acceptable length, it was decided to keep the texts short, but complete and cohesive, i. e. not marked as the excerpts they actually were. Some CTS researchers argue that the first minutes into a new task is when important decisions for the whole text are taken and some task familiarisation is still taking place (Muñoz and Cardona 2018, 23). After this initial contextualisation, the translator may resort to her routine behaviour and automatised processes (Göpferich 2013, 67). Having a short source text therefore may have an influence on the cognitive effort spent and may not be as representative of everyday tasks as a longer source text would be. Muñoz, for example, states that "the notion of translating a whole text without interruptions for research purposes tends to prompt researchers to use originals which are far shorter than the average text length in real commissions, and differences between long and short translation tasks still remain to be accounted for" (Muñoz 2010, 181–182). It was important to me to have recordings of complete processes and to analyse those rather than only a portion of a longer process. It is feasible to assume that a completed short task includes all types of activities that also usually occur in longer tasks. In addition, task fatigue probably plays a less important or no role in short tasks. Since it is not an object of study here, its potential absence has no bearing on the results. Participants came to the lab at times that suited them, and I assume that since they agreed to participate and none mentioned any inconvenience during the recording or in the interview that would have had a severe influence on the process and the product, all were more or less satisfied with how they had proceeded and what they had produced. The evaluation of those products was the focus of the substudy, as explained in the next section.

4.3 Substudy: Directionality and the translation product

Before translation processes were systematically observed and recorded in research studies, quality assessment was generally related only to the translation product. As this is still how it is commonly defined in many areas, I will also only consider here the assessment of the quality of the translation product and not of the process. Assessing the quality of the TT has often and mainly been based on notions of equivalence, which made the ST – and the evaluator's interpretation of it – and the evaluator's language skills two crucial determinants for evaluation. Since the rise of the functional approach in translation studies, the perceived target audience and the translation commission have also gained weight as factors to consider in product quality evaluation. Some TS researchers and trainers have gone a step further to consider the end-user's view in TT evaluation and therefore make it less dependent or completely independent of the source text (as discussed in Section 3.3). I also opted for this approach to design my substudy on the translation product.

The aim of this substudy is to gain information on the product and its relation to directionality as expressed in the research question 'In what ways are the products of professional L2 translation similar or different from the products of professional L1 translation?' Within the scope of this study, translation product refers to the final target texts (TTs) produced during the translation processes observed in the main study. There, in a first translation task, six German bidirectional translators (Bidir) and six English unidirectional translators (UniEnglish) produced an English TT each. In a second translation task, the same bidirectional translators and six German unidirectional translators (UniGerman) produced a German TT each. The resulting TTs represent the object of study of the substudy. In the following, I present the substudy's design, its participants, stimuli and tasks, data collection methods and data collection procedures. I also report on the operationalisation of the characteristics I investigate and discuss ethical and methodological issues.

4.3.1 Substudy design

In the substudy, the target texts produced by the bidirectional translators are compared to those produced by the unidirectional translators with the same target language. In order to do this, categories of text characteristics need to be established. Two surface linguistic features will be used to investigate their comparability in general: *word count* and *character count*.

As it is generally assumed that the quality of L2 translation is poorer than that of L1 translation, this is examined in the English target texts produced by the bidirectional translators into their L2 (Bidir_L2) and the English unidirectional translators into their L1 (UniEnglish_L1). Their quality is determined by analysing *readability* and *acceptability*, which is assessed in a ranking task. For the ranking task, it makes sense methodologically to have it performed by people not involved in the study otherwise. As the evaluation should also be in line with the requirement of ecological validity, I decided to have it performed by potential members of the intended target audience, which in this case are newspaper readers. This approach required the recruiting of raters and the development of a questionnaire, which is described in the next two sections *Substudy participants* and *Substudy stimuli and task*. An overview of the type of data elicited in the substudy is presented in Table 4.

4.3.2 Substudy participants: Text users as evaluators

Undoubtedly, the potential target audience is a factor that needs to be taken in consideration when performing a translation task. In the case of translation, the members of this target audience are often the readers of the target text. For the translator, this end-user is not a real person but a concept (Suojanen et al. 2015, 62–63). In the case of external translation evaluation, this reader becomes a real person, i. e. the evaluator. Colina supports choosing the type of evaluator according to the purpose and goal of the evaluation (2009, 255). As my main criterion was that the evaluators be suited to the evaluation task (Doherty 2017, 140) and to my research purposes, they needed to have a certain level of English language skills and be prepared to participate. As the user-centred approach would call for readers of newspapers for these TTs, I wanted non-professional evaluators.

The recruiting of participants for the product evaluation task was done via my research network. As I wanted not only raters with English as their L2, which were not particularly difficult to find in Switzerland, but also raters with English as their L1, I needed support from other researchers. As representatives of the intended target audience, meaning potential readers of a quality English-language newspaper, university students with different language backgrounds were chosen (see Table 3). There were two types of participant groups with 18 participants each. One group of raters had English as their L1 (L1 raters), which means that they rated the twelve English texts in their L1. The other group of raters had English as their L2 or Ln (L2 raters). In the following, the two groups are described in detail.

L1 raters

The group of L1 raters was recruited via a lecturer and translation process researcher at a liberal arts university in the US. Of the 35 students who participated in the study, 18 were randomly selected – using an online random number generator – so as to balance the size of the second group of raters. All L1 raters have English as their L1 and were raised monolingually. They were Bachelor's students in their first to third year, studying towards a degree in nursing (n=8), biology (n=2) and other disciplines (n=8) but not a language degree. They all participated in a basic Spanish language course. They were between 18 and 20 years old at the time of data collection, and 61 % were female (see Table 3).

	Group size	**Participants**	**L1**	**L2/Ln**	**Average age (years)**	**Female (%)**	**'Formal' English at age**	**Main area of study**
L1 raters	n=18	bachelor students (US)	English	Spanish	19	61	6	Nursing (8), Biology (2), other (8)
L2 raters	n=12	bachelor students (CH)	German	English	23	83	12	Applied Languages (12)
	n=6	bachelor or master students (GB)	neither English nor German	English	26	83	7.5	all different ones

Table 3: Rater groups in target text evaluation task

L2 raters

The L2 group of raters all have English as their L2 or Ln. Twelve of them were recruited via a lecturer and translation process researcher at the Zurich University of Applied Sciences. They were Bachelor's students in 'Angewandte Sprachen' (Applied Languages) at the beginning of their 3rd semester. They all stated that German[30] was their L1, and for all but one this was also the only language they spoke before entering school. On average, they had started to learn English at the age of twelve, that is in grade 7. They were between 20 and 31 years old, and 83 % were female. The other six members of this group were recruited via a lecturer and researcher at a public research university in Great Britain. They were all studying towards different degrees – four at Bachelor's (in their 2nd or 3rd year) and two at Master's level. As their L1, they mentioned Mandarin, Russian, Arabic, Spanish, Hungarian and Latvian,

30 10 of 12 participants stated that Swiss-German was their L1, which is treated as a dialect and not as a distinct language for the purposes of this study.

respectively. On average, they had started to learn English at school at the age of seven. They were between 19 and 43 years old, and 83 % were female.

As all participants were studying in English or English-taught courses at the university level, a sufficient level of language skills for the task can be assumed. Therefore, an English language proficiency test was not considered to be necessary for any of the rater groups. This assumption is also supported by the fact that none of the participants abandoned the task or commented on not being able to perform it.

4.3.3 Substudy stimuli and task: English target texts and their evaluation

The six members of the group of bidirectional translators produced one English target text each of between 105 words (Bidir6_L2_TT) and 133 words (Bidir1_L2_TT) with a median of 120 words. The six unidirectional translators into English also produced a TT each. Their lengths ranged from 113 words (UniEnglish1_L1_TT) to 136 words (UniEnglish2_L1_TT), also with a median of 120 words. Whereas all six bidirectional translators produced a different TT title for the German *Strandungen von Walen*, the unidirectional translators produced only four different versions, with *Beached whales* being the favourite solution (3 out of 6). All twelve TTs were complete with respect to the propositional content of the ST.

Data was collected from representatives of the intended users of the translated texts. With the help of a questionnaire, they were asked to rank the texts based on their acceptability for publication in a quality English-language newspaper. The evaluation criteria were completely their own. To further simplify matters, they only needed to decide on the three most acceptable and the three least acceptable texts out of the twelve.

As a preparation for the evaluation task, all twelve TTs were identically formatted as to prevent content-independent aspects such as typographical aspects to affect their assessment (Conde 2011, 75). Six sets of the twelve translations were prepared, each containing each text printed on a separate page and arranged in random order. The only constraints were that none of the texts appeared more than once in either first or last position or not more than twice in any other position. This was done in order to balance any order effects of serial translation evaluation (Muñoz and Conde 2007). Moreover, the texts were coded to allow for blind assessment by the evaluators. They did not have access to the source text.

The instructions to the evaluators were presented on the first page of the questionnaire and provided as much information as needed to accomplish the task at hand without revealing exactly how the TTs had been obtained or who had produced them (see Figure 4).

Ranking twelve journalistic texts

Dear participant
The following twelve texts were produced by professional translators. They had been asked to translate a short text for a quality English-language newspaper. You, as a potential reader of that newspaper, can now decide which of the twelve texts are the three most acceptable ones and which are the three least acceptable ones.
Please do not change the order of the texts. Please work on your own and do not look anything up but feel free to annotate the texts. After having decided on the ranking, please fill in the ranking sheet and the background information on the last two pages. All information will be treated anonymously and is only used for research purposes.

Figure 4: Instructions for evaluators of English TTs

4.3.4 Substudy data collection methods

In order to investigate the product descriptors related to length, all 24 target texts were used as data source. The data was collected by running the word and character count application within Microsoft Word for each TT. The product quality of the English TTs was determined by their readability and their acceptability. Data on the TTs' readability was collected by applying the Flesch Reading Ease formula to each of them. Data on the acceptability of the TTs was collected by a ranking task (see Table 4). The results were then transformed into scores for each TT (see Section 4.3.6).

Part of substudy	Type of information	Data collection instrument	Form of data	Type of data
Product descriptors	Word and character counts	word count in Microsoft Word	numbers	quantitative
Product quality	Readability of English target texts	Flesch Reading Ease formula	scores	quantitative
	Acceptability of English target texts	calculation based on results from ranking task	weighted scores	quantitative
	Personal background of raters	questionnaire	transcript	qualitative
	Raters' individual evaluation criteria	questionnaire	transcript	qualitative

Table 4: Overview of data collected in substudy

The ranking task was questionnaire-based and designed for members of the intended target audience. The participants were asked to evaluate the acceptability of the presented target texts by ranking the three most acceptable ones and the three least acceptable ones. They were also asked

about their criteria for evaluation. The questionnaire contained additional open and closed questions and was pilot-tested. On the first page, the participants were given instructions, informed about their rights and how to contact the researchers. Then followed the 12 TTs, each on a separate page. On page 14, the participants were asked to document their evaluation in a ranking list (see Figure 5).

Ranking

Please enter the codes of the texts you chose. You find the code on the same page as the respective text.

Rank	Document code	
1.		most acceptable
2.		
3.		
⋮	⋮	
10.		least acceptable
11.		
12.		

Figure 5: Template of ranking list for raters

Each of the six different sets of the twelve texts was preceded with instructions and followed by a ranking sheet, Likert scales for task difficulty, a question on chosen criteria for evaluation and questions in order to collect some background information on the participants. For a copy of the questionnaire see Appendix B.

4.3.5 Substudy data collection procedure

This section describes the procedure to collect data on the acceptability of the English TTs. As each of the rater groups (L1 raters from the US, L2 raters from Switzerland and L2 raters from Great Britain) was located in a different country, the preparations for data collection depended on the three involved universities' requirements for student participation and recruitment procedures.[31] In order to obtain approval from the foreign universities, on-site researchers had to be appointed co-investigators. On the occasion of the actual data collection, the respective on-site (co-)investigator distributed

31 An account of the different procedures for approval by the universities' research ethics committees is given in the section on ethical issues.

the questionnaires to the raters. They had 30 minutes to complete the task and were prompted after 25 minutes. The (co-)investigator then collected the questionnaires. The setting varied somewhat between the three locations due to the different formal procedures the data collectors had to follow at their respective institutions. At the US university, the L1 raters participated during class time and were rewarded with credits. At the British university, the L2 raters participated outside of class time at the co-investigator's office and were not awarded any credits. At the Swiss university, the L2 raters participated during class time without getting credits. The different circumstances under which data collection was possible had an impact on the number of participants in each rater group. Apart from that, it did not present any methodological issues.

The next section covers the operationalisation of the units of analysis in the product analysis and product evaluation parts of the substudy.

4.3.6 Substudy operationalisations

As already introduced in the section on the substudy design, the approach to insights into the quality of the translation product is a holistic one: it focuses on subjective judgements by potential readers and macro-textual features rather than a detailed linguistic analysis. Moreover, in keeping with the notion of descriptors adopted for the translation process in the main study, descriptors for the translation product are introduced and operationalised in the following sections.

Product descriptors

In the following, the product descriptors for each category of text characteristics are defined. They cover the aspects of the translation product investigated in this study.

Number of TT characters is defined as the total number of characters and spaces present in the final TT, including the title. This descriptor is used for all 24 target texts produced in the main study, that is Bidir_L2_TTs, UniEnglish_L1_TTs, Bidir_L1_TTs, and UniGerman_L1_TTs.

Number of TT words is defined as the total number of words present in the final TT, including the title. This descriptor is also used for all 24 target texts produced in the main study.

Productivity

Productivity is the number of words in the translator's final target text that she produces per unit of process time. It is calculated as the number of TT words divided by process duration in minutes and reported as words per minute.

Product quality

As the primary focus of this study is on L2 translation into English, product quality is exclusively applied to the English TTs. Product quality consists of readability and acceptability.

Readability is defined here as the Flesch Reading Ease score. Readability refers to textual organisation such as sentence length, paragraph division, word choice, argumentation structure, rhythm and coherence (Suojanen et al. 2015, 49–50). The Flesch Reading Ease score is one of the readability indices that are based on sentence length and word length. The idea behind the concept of these readability indices is that short sentences and short words are easier to understand than long ones. Flesch (n. d.) inversely related the scores, which can range from 0 to 100, to grade levels within the US school system or, in other words, the higher the score, the easier a text is to read and hopefully also to understand. As Suojanen et al. (2015, 52) and Si and Callan (2001) point out, however, these formulas do not say anything about cohesion or whether the text makes sense at all. Moreover, they should only be used in the languages they were developed for, which is English in most cases. All the same, they offer a quick and straightforward indication of the complexity of a text based on its structure.

Acceptability is defined as a combination of weighted scores for every position in the ranking task. As there are twelve target texts to be evaluated, there are 12 possible positions. However, in order to reduce task complexity, the raters only have to determine positions 1 to 3, the top three ranks, and positions 10 to 12, the bottom three ranks.

position	score
1	12
2	11
3	10
4 to 9	0
10	-10
11	-11
12	12

Table 5: Scores assigned to the positions in the ranking task to determine product acceptability

As the positions 4 to 9 are excluded from the ranking task, they do not receive a score. Assuming that those positions indicate a medium level of acceptability compared to the other positions (and TTs), they are considered the neutral baseline or starting point (i. e. zero) for the product quality score. For this reason, the scores for the top 3 positions are treated as bonuses and therefore added while those for the bottom 3 positions are treated as penalties and therefore subtracted from the starting score of 0. The scores are listed in Table 5. As there are 36 raters to rank the TTs, the maximum possible score is 432 and the minimum is -432.

Raters' evaluation criteria

In the evaluation task, participants were asked to list the criteria they had used to assess and rank the target texts' suitability for publication. These criteria are then categorised, counted and the results compared between rater groups.

4.3.7 Substudy ethical issues: Students as participants

For many researchers who work at educational institutions such as universities, students as participants may be the obvious and first choice. To make sure that the students' rights as dependent and therefore potentially vulnerable participants are not infringed upon, many research institutions have introduced guidelines and standards. Requirements differ from country to country, and even from university to university. This was certainly the case in my questionnaire-based product evaluation that involved participants from the Zurich University of Applied Sciences in Switzerland, a liberal arts university in the US and a public research university in Great Britain. In Switzerland, most universities have internal research boards that issue recommendations and good practice information. At ZHAW, where I collected data from L2 raters, no approval by the cantonal ethics committee was required. In consultation and agreement with the students' instructor, I decided on an ethical procedure in which the students participated anonymously and were informed about their right to withdraw during or after the data collection. Moreover, they were assured that their participation or non-participation would have no bearing on their grades.

The other student rater groups also participated under these terms and conditions. As a prerequisite to conducting a study with students at the US university, I had to take and pass an online course on research ethics provided by the Collaborative Institutional Training Initiative. It consisted of ten modules on topics such as the US federal regulations, assessing risk, informed consent, privacy and confidentiality. A fellow researcher

who taught at the institution then had to submit a comprehensive set of documentation to the ethics committee, stating that he would be the principal investigator for that part of the study. Approval by the research ethics committee responsible also had to be obtained for the British university. They required explicit informed consent and a signed form whereas the US committee suggested a completely anonymous procedure. Although the procedures and requirements for conducting studies with student participants may differ, the main requirement remains: the students' rights and integrity shall not be infringed upon.

4.3.8 Substudy methodological issues

Within Translation Studies, product evaluation has been approached from diverse angles (for a discussion relevant for this study see Section 3.3). If evaluation is performed by humans, and not algorithms, "[t]he essential aspect [...] is, of course, the evaluator, whose skills and attributes must be suited to the evaluation task" (Doherty 2017, 140). In my study, the two rater groups for the ranking task, the L1 raters who read the texts in their L1 and the L2 raters who read them in their L2, need a certain level of English skills in order to perform the task as required. Therefore, I chose university students as participants. This also makes them comparable with regard to their level of education. In the interest of transparency and rigorous documentation, detailed information of evaluators and evaluation is provided in Sections 4.3.2 and 4.3.5 and in Appendix B.

4.4 Directionality and translator effort

As explained in Section 3.4, translator effort is a construct that can only be measured indirectly and according to the framework of situated and embodied cognition cannot be meaningfully separated into different types of effort. Translator effort is defined as the amount of exertion a translator invests into the execution of a translation task. This includes everything from the muscle tension needed to remain seated during the TP to pressing and releasing keys on the keyboard in the final revision and proofreading phase. As all these observational and other measures are indirect indicators of translator effort, several of them need to be analysed in order to get a representative picture of translator effort.

Task execution demands effort on the part of the task executor, the translator. She is mainly the one who decides how much effort she puts into task completion, although this may not be a conscious decision. As translating is principally a cognitive activity and cognition involves

permanent perception of and interaction with the environment, there is no such thing as a 'mindless' activity. Therefore, every observable activity is an indicator of underlying cognitive activity. This is under the premise that the translator actually sets her mind to translating. The relationship between observed activity and underlying cognitive processes is not straightforward and is always open to interpretation. For example, fast typing is not necessarily indicative of high cognitive load but neither is slow typing or pausing. However, some indicators seem to be better indicators of cognitive load than others, especially if analysed over a whole translation process. A triangulation of process and product descriptors forms the basis for these indicators (see Table 6). As with many complex cognitive activities, a single indicator in a TP does not yield enough information in isolation, so an array of indicators is proposed. They need to have internal validity, which I interpret as being closely connected to the visible activities performed in the TP. This section is devoted to the operationalisation of such indicators of translator effort.

Indicator	Definition	Process measure(s) used	Product measure used
Character production effort	Ratio of number of characters in process to number of characters in final product	Character count	Character count
Revision effort	Number of revisions per 10 TT words	Total number of revisions	Word count
Information search effort	Number of search queries per 10 TT words	Total number of information searches	Word count
Coordination effort	Percentage of pauses with gaze focus on ST and TT	Total number of pauses. Representation of gaze data in SCR	n.a.
Mean fixation duration	Mean fixation duration during whole TP	Mean fixation duration	n.a.

Table 6: Indicators of translator effort and their components

An ideal translation process (i. e. one with the least effort involved) would look roughly as follows: the translator types the target text while reading the ST all in one go. Of course, this is usually completely unrealistic, since most texts are complex and translation requires work. All the same, this is the starting point for my choice of indicators of translator effort. As stated in the literature review, I do not differentiate between indirect and direct indicators of translator effort nor do I use different types of effort (for an example of a different approach, see Alves et al. 2016). I take the stance that based on the paradigm of situated and embodied cognition, all indicators of translator effort are at the same time indicators of cognitive effort, as human cognition cannot be separated from the human body that expends

this effort. The translator effort measures in my study mostly comprise triangulated process and product data. This renders them ST-independent and reflects the complexity of the translation task.

4.4.1 Character production effort

Character production effort is defined as the ratio of number of characters and spaces produced in the process (i. e. a process measure) to the number of characters in the final TT (i. e. a product measure). In other words, this provides an indication of how many characters were needed to produce one character of the TT.

4.4.2 Revision effort

Here, *revision effort* refers to the number of revisions performed during the whole translation process relative to the number of words present in the final TT version. As revision is usually word based, normalisation per TT word makes sense.

4.4.3 Information search effort

Information search effort is an indicator that relates the total number of search queries, which are either performed on their own (as single-step searches) or in clusters (as part of multi-step searches) during the translation process to the number of words in the TT.

4.4.4 Coordination effort

As I mentioned before, I had refrained from analysing reading activities in the TP so far. In order to acknowledge its importance, I decided to integrate it into one of the effort indicators: *Coordination effort* is based on the assumption that the TT needs to be aligned with the ST with respect to completeness of the translation and interpretation of the content and mapping to the purpose of the text. In other words, this is done by reading and comparing the ST and the TT. This switching of attention between the TT and the ST can, for example, be done during long pauses, which here are operationalised as the process descriptor *pausing* and last longer than five seconds. I therefore determine the proportion of pauses to which this applies based on the visualisation of gaze data in the screen recordings. For a

pause to fall under the category of focusing on TT and ST, the visualisation of the translator's eye movements needs to fall at least once on the ST and once on the TT within that pause. In most cases of this category, the translator is actually switching back and forth between the ST and the TT. Coordination effort may, on the one hand, simply depend on the individual translator's preferences in proceeding. On the other hand, it may be related to directionality, assuming that L1 translators – who read the ST in their L2 – need to coordinate more often between ST and TT than L2 translators do.

4.4.5 Mean fixation duration

Mean fixation duration is defined as the average length of the total number of visual fixations over the whole translation process. To measure visual fixations, an eye-tracking device is used (see Section 4.2.4 for details).

4.5 Relations between characteristics of professional translation

The process and product measures triangulated in the forms of productivity and translator effort lend themselves to further correlation analyses with product quality and the concept of *professional experience*. The variable professional experience was elicited in the initial interview and is part of participant data as illustrated in the overall study design in Figure 1. I operationalise professional experience as the number of years after completion of a degree that a person has worked as a professional translator. I do not distinguish between those who work full-time and those who work part-time. Professional experience is based on the self-report data that I collected during the interviews at the ZHAW usability lab. Questions that I address in my analyses of additional relations include: Do more experienced translators expend more effort than less experienced translators do? Is high translator effort related to high product quality? Are there differences with respect to translation direction?

Although it may be tempting to infer causality based on strong correlations between variables, results of correlation analysis could also be due to chance or indicate an indirect relation via a third variable or a reciprocal connection. In order to investigate any causality or predictability between the indicators, larger participant groups and appropriate statistical tests such as linear regression modelling would be needed.

The next chapter presents the results of the main study, that is the investigation into the translation processes by means of the process activities writing, revising, information seeking and pausing.

5 The translators' processes

In this study, I investigate the translation process (TP) with the help of self-defined process descriptors, two global ones and four that are related to the main process activities. I chose this approach rather than fragmenting the TP into arbitrary micro units. All analyses and measures are produced with the goal of addressing my research question 'In what ways are the processes of professional L2 translators similar to or different from the processes of professional L1 translators?' The results of the bidirectional translators' processes into their L2 and into their L1 (i. e. Bidir_L2 versus Bidir_L1) are presented first for each measure and analysis. This allows for intragroup comparisons in order to identify potential patterns of individual behaviour in both translation directions.

The intragroup comparisons are followed in each case by the two intergroup comparisons: the results for the Bidir_L2 group are compared to those of the UniEnglish_L1 group as they both performed the same German-English task, and the results for the Bidir_L1 group are compared to those of the UniGerman_L1 group as they both performed the same English-German task. This approach should help identify potential similarities and differences that might be related to directionality. Some of the process parameters are later triangulated with product parameters to act as indicators of translator effort.

The dependent variables in this main study are the six process descriptors derived from KSL, SCR and RVP data as defined in Section 4.2.6 in the *Methodology* chapter. Descriptive statistics and inferential statistical tests were used wherever justifiable. Since the sample size is small and normal distribution of the data cannot be assumed, non-parametric statistical tests are used for all results in this chapter. For intragroup comparisons, the Wilcoxon signed-rank matched-pairs test is used.[32] For intergroup comparisons, Welch's *t*-test on the ranks of the data is used as appropriate.[33] The level of significance is 5 % unless otherwise indicated. All results are reported on. However, while all of the results for the intragroup comparisons are illustrated with a table or graph (except for *total number*), those for the intergroup comparisons are only presented in detail if there were any statistically significant differences. In the interest of transparency

32 I used the calculator supplied by Mellinger and Hanson (2016), and as recommended there, report test statistic (T), approximate p-value and effect size (r).

33 I used XLSTAT, version 2018.2 (XLSTAT 2017). I performed two-tailed tests and report observed value of the t statistic (t), degree of freedom, p-value (p) and confidence interval (CI).

and rigorous documentation, all of the comparisons are presented in tables and figures in Appendix C.

5.1 Global process measures

In order to provide an initial description of the processes in this study, two global measures are presented: 1) the process duration, which equals the time the translators spent to produce the translation, i. e. the final target texts; and 2) the number of characters and spaces that the translators produced during the translation process TP (for the definition of TP see 4.2.6 in *Methodology*). Both measures were extracted from the KSL data files as the main data source and are described in more detail below.

5.1.1 Process duration

The execution of the translation task, that is the production of a final TT, takes a certain amount of time, which is measured in seconds and referred to as process duration in this study. To start with, the results for the six bidirectional translators' processes into L2 (Bidir1_L2 -Bidir6_L2) and the same translators' processes into L1 (Bidir1_L1-Bidir6_L1) are presented (see Table 7). Process duration for Bidir_L2 ranges from 1335 seconds to 2842 seconds and the median is 1996 seconds. The Bidir_L1 process duration ranges from 1136 seconds to 2085 seconds and the median is 1618 seconds. Five of the six translators took longer translating into their L2 than into their L1. However, there is no statistically significant[34] difference between the translation directions with regard to process duration.[35]

Process	Process duration (sec)	Process	Process duration (sec)
Bidir1_L2	1586	Bidir1_L1	1136
Bidir2_L2	1708	Bidir2_L1	1649
Bidir3_L2	2842	Bidir3_L1	1545
Bidir4_L2	1335	Bidir4_L1	1586
Bidir5_L2	2283	Bidir5_L1	2049
Bidir6_L2	2510	Bidir6_L1	2085

Table 7: Duration of Bidir_L2 and Bidir_L1 translation processes (in seconds) ordered by translator

34 I use the term *significant* only in the sense of *statistically significant* here and throughout.

35 Wilcoxon signed-ranks matched-pairs test: T=3; approximate p=0.12; r=0.71.

In the following, process duration is compared between groups. First, the results for the Bidir_L2 group are compared to those for UniEnglish_L1 group, both of whom translated the same German ST into English. Process duration for UniEnglish_L1 ranges from 640 seconds to 1407 seconds and the median is 1199 seconds. Table 8 shows the process durations for the two groups ordered from fastest to slowest. Five of the six UniEnglish_L1 processes are shorter than those of the Bidir_L2. In fact, there is a statistically significant difference in process duration between the two groups.[36]

Process	Process duration (sec)	Process	Process duration (sec)
UniEnglish1_L1	640	Bidir4_L2	1335
UniEnglish5_L1	922	Bidir1_L2	1586
UniEnglish6_L1	1081	Bidir2_L2	1708
UniEnglish2_L1	1318	Bidir5_L2	2283
UniEnglish3_L1	1396	Bidir6_L2	2510
UniEnglish4_L1	1407	Bidir3_L2	2842

Table 8: Duration of UniEnglish_L1 and Bidir_L2 processes in ascending order (in seconds)

The second intergroup comparison is between the Bidir_L1 and UniGerman_L1 group who translated the same English ST into German. Again, the results for process duration are ordered from fastest to slowest (see Table 9). Process duration for UniGerman_L1 ranges from 887 seconds to 1896 seconds and the median is 1141 seconds. A comparison shows that the results only overlap partially. This finding is corroborated by a Welch's test on the ranks of the data that indicates a significant difference between the two groups in terms of process duration.[37]

In summary, the bidirectional translators took significantly longer to produce their translations than the unidirectional translators did, irrespective of translation direction. Subsequent analyses with other measures that are more detailed provide more insights into the potential reasons for this. The next section presents the results for the second global measure, the number of characters produced during the TP.

36 Welch's *t*-test conducted on the ranks of the data: t[9.95]=3.85; p=0.003; 95 % CI [1.77, 8.90].

37 Welch's *t*-test conducted on the ranks of the data: t[9.84]=2.25; p=0.049; 95 % CI [-0.57, 8.57].

Process	Process duration (sec)	Process	Process duration (sec)
UniGerman4_L1	887	Bidir1_L1	1136
UniGerman2_L1	1021	Bidir3_L1	1545
UniGerman6_L1	1126	Bidir4_L1	1586
UniGerman3_L1	1157	Bidir2_L1	1649
UniGerman5_L1	1165	Bidir5_L1	2049
UniGerman1_L1	1896	Bidir6_L1	2085

Table 9: Duration of UniGerman_L1 and Bidir_L1 processes in ascending order (in seconds)

5.1.2 Character count

The translators produce characters (i. e. letters, numbers, punctuation marks) and blank spaces when they use their keyboard in the TP. They enter them while writing TT, revising TT (= inserting text) and initiating information searches. All of these characters and spaced were included in *character count* in order to give a first impression about the amount of effort expended in the TPs.

Process	Character count	Process	Character count
Bidir1_L2	1185	Bidir1_L1	990
Bidir2_L2	1159	Bidir2_L1	1073
Bidir3_L2	1507	Bidir3_L1	1224
Bidir4_L2	1103	Bidir4_L1	1191
Bidir5_L2	1401	Bidir5_L1	1333
Bidir6_L2	1851	Bidir6_L1	1426

Table 10: Character count for Bidir_L2 and Bidir_L1 processes (no.) ordered by translator

In the following, the results for the TPs of the bidirectional group translating into their L2 (i. e. Bidir_L2) are compared to those when they translated into their L1 (i. e. Bidir_L1; see Table 10). The aim is again to detect potential patterns on an individual level, that is to see whether those translators who produce relatively low character counts when translating into their L2 also have similar ones when translating into their L1. The results for the Bidir_L2 processes range from 1103 characters to 1851 characters and the median is 1293 characters. For the Bidir_L1 processes, 990 characters to 1426 characters were produced, with a median of 1208 characters. There seems to be a tendency that the processes into L2 with a low character count also have a low character count into L1, as those ranked in positions 1 to 3 in the Bidir_L2 processes also rank in positions 1 to 3 in Bidir_L1 processes. This

could point to an individual behaviour of the translators: While all have to type the characters that ultimately result in the TT, some may tend to do a lot of inserting and/or to use long search strings in information seeking, while others do not. Indeed, the statistical test showed that there was no indication of a significant difference between the L2 and L1 processes for the bidirectional translators with respect to character count.[38]

The following analyses compare the results for the unidirectional translators' processes with those of the bidirectional group with the same tasks to explore whether character count may be related to directionality. Results of the character count for the UniEnglish_L1 group range from 829 characters to 1381 characters, and the median is 1100 characters. Compared to the Bidir_L2, there seems to be a tendency for the UniEnglish_L1 group to produce fewer characters during their TPs. However, there is no significant difference between the two groups.[39]

The results of the character count for the UniGerman_L1 group range from 859 characters to 1397 characters and the median is 1098 characters. Compared to the counts for the Bidir_L1 group, there seems to be some similarity between the groups, and in fact, there is no significant difference between the two groups (see Appendix C for individual results).[40]

In summary, whether a bidirectional translator tends to produce a high or a low number of characters seems to be similar in both translation directions, which may point to an individual pattern of text production. When compared between groups, the number of characters produced is not significantly different. Therefore, character count does not seem to be related to directionality.

5.1.3 Summary of results for global process measures

This section has shown that there is a significant difference between the bidirectional and the unidirectional translators' processes with regard to process duration, but no apparent effect of directionality as the bidirectional translators have longer process durations in both translation directions. It also showed that there is no significant difference between the bidirectional and the unidirectional translators' processes with regard to the total number of produced characters. While the amount of typing has some influence on process duration, it does not seem to be the only contributing factor.

38 Wilcoxon signed-ranks matched-pairs test: T=3; approximate p=0.12; r=0.71.

39 Welch's *t*-test conducted on the ranks of the data: t[10.00]=2.13; p=0.059; 95 % CI [-0.79, 8.45].

40 Welch's *t*-test conducted on the ranks of the data: t[9.84]=1.13; p=0.28; 95 % CI [-2.95, 7.62].

These global measures provide an overall impression of the TPs and are later used in triangulation with other measures to derive additional indicators of professional translation (see Chapter 8). In isolation, they do not provide any insight into which process activities contributed to the results or into further similarities or differences between the groups. Therefore, finer-grained process units (i.e. the four process descriptors *writing*, *revising*, *information seeking* and *pausing*) are investigated in the following sections. To start with, the focus is still on the whole TP, which can be defined as a string of these four process activities.

5.2 General description of the translation processes

In all 24 processes that were recorded for this study, the translators produced their TTs in a similar way, starting with the title. They all produced an initial version of the TT that they then revised in several cycles. All TPs were transcribed based on the definitions for the four process activities used in this study (see 4.2.6 in Chapter *Methodology* for operationalisations and Appendix C for data preparation). For illustration purposes, two TPs are visualised as strings of those four activities (see Figure 6). They are the processes of the bidirectional translator *Bidir2* into her L2 and into her L1 (Bidir2_L2 and Bidir2_L1, respectively). They are illustrative of the other bidirectional translators' processes in terms of occurrence and distribution of activities as well as relative process durations.

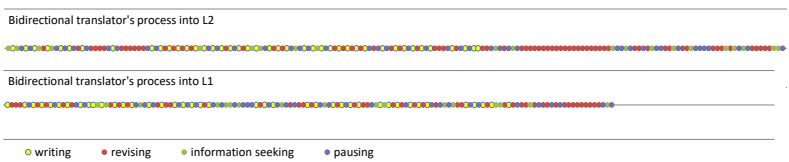


Figure 6: Examples of a bidirectional translator's processes into L2 and L1 (Bidir2_L2 and Bidir2_L1) as sequences of the four main process activities

Each dot of the strings in Figure 6 represents one of the four activities writing, revising, information seeking and pausing performed during the TP. Starting on the left, the dots are aligned in chronological order according to when the translator performed them during her TP. Whereas by definition a writing sequence is delimited by sequences of a different activity type, the other three types of activities can appear in clusters. In total, the bidirectional translator Bidir2 performed 164 activities within her L2 translation task and 137 activities within her L1 translation task despite the similar duration of the tasks (28.5 minutes and 27.5 minutes, respectively). The Bidir_L2 process starts with an information-seeking

incident followed by the first writing activity. For a long stretch during the first two thirds of the process, it is mostly writing that alternates with the other three activities. Then writing ceases as the first version of the TT is completed, and revising dominates during the last third of the process. It is noteworthy that both information seeking and pausing occur throughout the TP, which is a typical feature of the TPs in my sample. Another typical feature of the TPs in my sample is that revising incidents are often performed in clusters towards the end of the process. In the Bidir_L1 process, writing only ceases after about four fifths of the TP. Although in both TPs, the first writing incident is not preceded by a pause, the ST or part of it must of course have been read beforehand. However, the reading activity apparently was not long enough to be categorised as a pausing incident. As illustrated, transcribing and visualising the TP as a string of activities allows for counts of activities and for investigating potential patterns. In the following, these kinds of analyses are performed in order to provide information about differences and similarities between the translator groups.

5.2.1 Total number of process activities

To start with, the total number of all four categories of process activities that are used in this study is calculated for each of the 24 TPs (for a table with the detailed results for all processes see Appendix C). For the bidirectional translators' processes into their L2 (Bidir_L2), the total numbers of activities range from 104 to 202 and the median is 173. For the same translators' processes into their L1 (Bidir_L1), the total numbers of activities range from 107 to 227 and the median is 147. Between Bidir_L2 and Bidir_L1 processes, there is no significant difference with regard to the total number of activities.[41] Three of the six bidirectional translators performed a higher number of activities when translating into their L2 than when translating into their L1. Those translators who ranked high within the group in terms of numbers of activities in the L2 translation task did not necessarily do so in the L1 translation tasks.

For the unidirectional English translators' processes into their L1 (UniEnglish_L1), the total numbers of activities range from 89 to 197 and the median is 133. When the results for the UniEnglish_L1 are compared to those for the Bidir_L2 as the processes are based on the same source text, there is no significant difference with regard to number of process activities.[42] For the unidirectional German translators' processes into their

41 Wilcoxon signed-ranks matched-pairs test: T=9; approximate p=0.75; r=0.14.

42 Welch's *t*-test conducted on the ranks of the data for Bidir_L2 to UniEnglish_L1 comparison: t[10]=1.53; p=0.16; 95 % CI [-2.06, 8.06].

L1 (UniGerman_L1), the number of activities range from 88 to 201 and the median is 140. These results seem very similar to those of the Bidir_L1 processes, which are based on the same source text. In fact, a statistical test on the ranks of the data shows no significant difference between the two groups.[43]

In summary, there are no significant differences in total number of process activities during the TP within or between the groups. This suggests that the number of process activities is not related to directionality. The sequence of activities could also be described as switches between different types of tasks, which have been considered taxing for cognitive resources (Hvelplund 2017b, 79 and Whyatt 2018, 66).

5.2.2 Proportion of activity types in the translation processes

In order to get an additional general impression of the TPs in this study, the share of the different process activity types is calculated with relation to the whole TP. To start with, their proportion is calculated based on the number of their occurrences in each of the TPs. Figure 7 top left shows the results for the bidirectional translators' processes into their L2. As could be expected, all types of activities are represented in the Bidir_L2 processes. The range for *writing* is from 19 % to 34 %.[44] *Revising* comprises the largest proportion in all but one of the Bidir_L2 processes, starting at 28 % and ending at 45 %. *Information seeking* contributes between 9 % and 19 % to the total number of TP activities, thereby representing the smallest proportion. *Pausing* starts at 12 % and reaches 26 %.

In the bidirectional group's processes into L1 (Bidir_L1; see top right of Figure 7), all types of activities are represented as well. The results for *writing* range from 23 % to 31 %. *Revising* comprises the largest proportion in Bidir_L1 starting at 32 % and reaching 53 %. *Information seeking* contributes between 8 % and 21 % to the total number of TP activities, thereby representing the smallest percentage in all Bidir_L1 processes. *Pausing* ranges from 9 % to 29 %. When the results for the Bidir_L1 and the Bidir_L2 processes are compared, as they were produced by the same translators, there is a tendency that the proportions of the different types of activities remain rather stable, irrespective of translation direction.

43 Welch's *t*-test conducted on the ranks of the data for Bidir_L1 to UniGerman_L1 comparison: t[9.87]=0.62; p=0.55; 95 % CI [-4.17, 6.84].

44 The proportions have been rounded. Therefore, the total can vary +/-1 %.

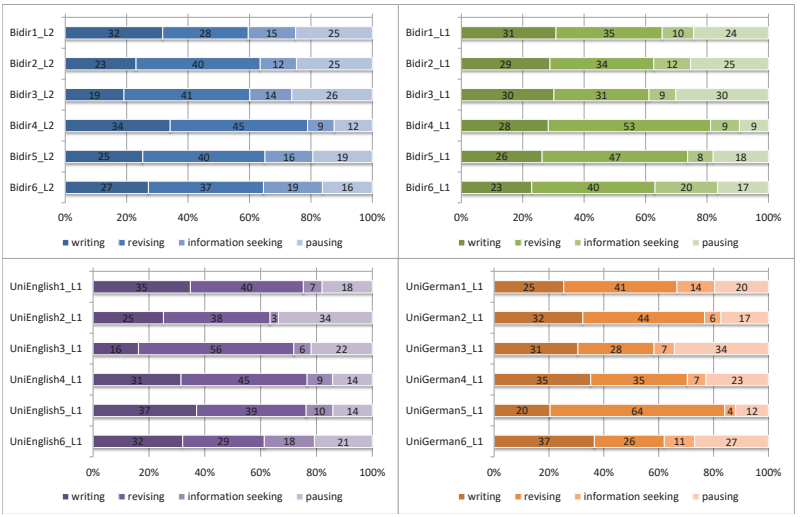


Figure 7: Proportions for no. of activity types in Bidir_L2, Bidir_L1, UniEnglish_L1 and UniGerman_L1 processes (%)

In the UniEnglish_L1 group (Figure 7, bottom left), the proportion of writing activities range from 16 % to 37 % of the total number of TP activities. Revising comprises the largest proportion in all but one TP, starting at 29 % and reaching 56 %. Information seeking contributes between 3 % and 18 % to the total number of TP activities, thereby representing the smallest proportion. Pausing starts at 14 % and reaches 34 %. Three of the six UniEnglish_L1 processes comprise rather similar proportions per activity types. However, this may be a coincidence. When the results of the UniEnglish_L1 group are compared to those of the Bidir_L2 – as these processes are based on the same ST – information seeking is the category with the lowest proportion in all TPs. Both groups share the tendency for revising to have the largest portion in the TPs based on number of occurrences, which emphasises the importance of this activity in the task. Overall, there do not seem to be large differences between the two groups.

For the UniGerman_L1 group (Figure 7, bottom right), the results for writing activities range from 20 % to 37 % of the total. Revising comprises the largest proportion only in three of the six processes, starting at 26 % and reaching 64 %. Information seeking contributes between 4 % and 14 % to the total number of TP activities, thereby representing the smallest fraction in all of this group's processes. The results for pausing range from 12 % to 34 %. All six of the UniGerman_L1 processes vary considerably with regard to the activity type proportions. A comparison of the results for the UniGerman_L1 group to those for the Bidir_L1, whose processes are based on the same ST, shows that information seeking is the category with

the lowest proportion in all but one of the TPs. Otherwise, no similarity between the two groups emerges.

This section has reported on the proportions that each activity type accounts for in the TPs based on number of occurrences. All four types of activities are present in all 24 of the TPs, with information seeking generally accounting for the lowest proportion and revising tending to account for the highest proportion. In the following, the proportions of the same activities within the TPs with respect to their relative durations rather than their numbers are considered and comparisons drawn within and between the groups.

As explained above, the four categories I defined for the transcription of the TPs cover the main activities performed during the TP. Their durations are calculated based on their definitions (see 4.2.6 *Methodology* Chapter). Apart from the main activities, other process activities can be observed. They include source text handling, TT formatting, highlighting text, moving within the TT document and between windows, scrolling in online resources, making short pauses (i. e. of five seconds or less), etc. In order to account for the total process duration, the time the translator spends on these activities is included in the following considerations and subsumed under *other activities*. The results are again reported for the Bidir_L1, Bidir_L2, UniEnglish_L1 and UniGerman_L1 but in a more general fashion than in the previous section. In all four groups (Figure 8), it is evident that the proportions of activities are distributed differently when based on their duration instead of their number.[45] Of course, this is also due to the integration of the category other activities that subsumes those activities not covered with the four main types writing, revising, information seeking and pausing. Of the total process duration, *other activities* accounts for proportions ranging from 18.3 % (UniGerman6_L1) to 50.3 % (UniGerman5_L1) reported over all groups. The biggest changes in proportions of the four main process activities occur in the categories revising and information seeking. While the former accounted for a large part and the latter for a small part in the TPs when based on number of occurrences, it is now the other way around. It is also noteworthy that the actual production of the TT – that is the aggregation of the two activities writing and revising – accounts for less than 25 % of total process duration in 22 of the 24 processes analysed. The time used for pauses from keyboard and mouse activities of longer than five seconds accounts for 8.9 % (Bidir4_L1) to 38.6 % (UniGerman6_L1) of the process duration. It is presumably during pauses when reading happens, for example of the translation brief, the TT, the ST or switching between TT and ST as well as cognitive activities

45 The proportions have been rounded. Therefore, the total can vary +/-1 %.

such as reflection, consideration of alternatives, etc. As a reminder: pauses during information searches are excluded from the analysis.

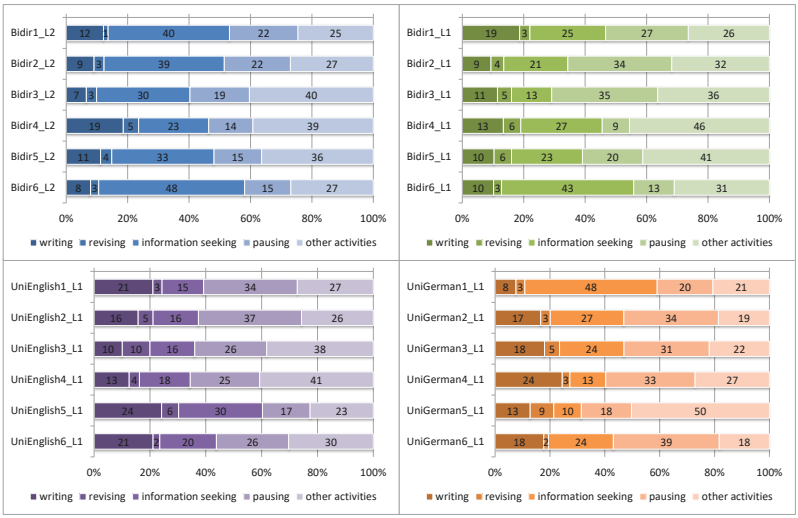


Figure 8: Proportions for duration of activity types in Bidir_L2, Bidir_L1, UniEnglish_L1 and UniGerman_L1 processes (%)

The comparison of the results for Bidir_L2 and Bidir_L1 (Figure 8, top left and top right, respectively) shows a tendency for the proportions of activity durations to remain consistent for both translation directions for the individual translators. This could be an indication of individual patterns in translating. When comparing Bidir_L2 to UniEnglish_L1 (Figure 8, top left and bottom left), there is a conspicuous difference in information seeking, which has a larger share in the TPs of Bidir_L2 than in those of UniEnglish_L1 at the expense of writing and pausing. When comparing Bidir_L1 to UniGerman_L1 (Figure 8 top right and bottom right), there does not seem to be a notable similarity other than that the variation within each of the two groups is rather large.

As for the activity type revising, it seems that the short amount of time needed to execute revisions is hardly related to the cognitive effort involved. My definition of revising is based on activities that are observable on screen and logged with KSL. However, the cognitive activity of revising may of course have started before that (e.g. when rereading ST or highlighting a TT segment) and continued after that (e.g. by monitoring the output). In other words, its perceived duration cannot be linked one-to-one to the underlying cognitive processes. These considerations apply to all types of process activities that I identified in this study.

5.2.3 Summary of results for the general translation processes measures

This section has considered the total number of process activities performed during the TP, which does not seem to be related to directionality. Moreover, the total number does not seem to indicate an individual pattern. It may rather reflect the translators' reaction to the task at hand. Moreover, this section has also considered the proportions that each activity type accounts for in the TPs with respect to their duration. It showed that while certain differences emerge between the Bidir_L2 and UniEnglish_L1 groups, there are none between the Bidir_L1 and the UniGerman_L1 groups. While this outcome may be related to directionality, this interpretation is challenged by the tendency of the bidirectional translators to behave similarly in both tasks. In order to delve deeper into the matter, the next sections examine each of the four main process activities in detail separately. The total number of occurrences and relative frequency are calculated for every activity, supplemented with other measures where appropriate. The process activities writing and revising are dealt with first, followed by information seeking and pausing.

5.3 Writing in the translation process

As defined in Section 4.2.6, the process descriptor *writing* applies to the activity when the translator adds new text at the end of the text produced so far. In general, the first writing incident of a TP starts with typing the first letter of the TT and the last writing incident ends with typing the last character of a first version of the TT. In the following, some aspects of the descriptor *writing* are investigated. The first is the total number of writing activities per process. The results are again presented first for the processes of the bidirectional translators into their L2 and into their L1 (Bidir_L2 and Bidir_L1), and these are then compared to the results for the unidirectional translators' processes into their L1 English or German, respectively (UniEnglish_L1 and UniGerman_L1). All comparisons are shown in Appendix C.

5.3.1 Total number of writing incidents

In the processes of the bidirectional translators working into their L2, the number for all *writing* incidents ranges from 33 to 55, and the median is 44 (for individual numbers of all 24 processes see Appendix C). In the processes of the bidirectional translators into their L1, the range is from 33

to 60, with a median of 40. An intragroup comparison shows that there is little variation in range and median of writing incidents. This is supported by the statistical analysis, which shows no significant difference between directions.[46] When the same translators' processes into L2 are compared to those into L1, four of them show remarkable consistency in their number of writing incidents for both translation directions.

In order to ensure that the measure is not source text dependent, it is also taken for the unidirectional translator groups. For the unidirectional translators working into English (UniEnglish_L1), the total number of writing incidents ranges from 26 to 62 and the median is 34. An intergroup comparison between the results for the UniEnglish_L1 processes and those for the Bidir_L2 processes, which are based on the same source text, shows that there is no significant difference with regard to the number of writing incidents.[47]

In the processes of the unidirectional translators working into German (UniGerman_L1), the total number of writing incidents ranges from 31 to 54 and the median is 42. A comparison of the results for the UniGerman_L1 and those for the Bidir_L1 processes, which are based on the same source text, does not yield a significant difference.[48]

In summary, there are no significant differences in total number of writing incidents during the TP between the groups who translated the same STs. This suggests that the number of writing sequences that translators perform during TT production is not related to directionality. The next section is concerned with the frequency of writing.

5.3.2 Frequency of writing

The frequency of writing, which is a measure of the number of activities per unit of time, indicates whether the rate of producing new TT remains consistent for both translation directions (Bidir_L2 compared to Bidir_L1). A comparison between groups with the same source texts (Bidir_L2-UniEnglish_L1 and Bidir_L1-UniGerman_L1, respectively) will show whether the large differences in process duration and the similarity in number of writing incidents resulted in a difference in frequency or not. Frequency of writing is defined as the number of writing incidents per ten minutes of process time. Within the Bidir_L2 processes, the frequency of writing ranges from 7 to 25 writing incidents per ten minutes, and the median is 13 writing incidents per ten minutes. Within the Bidir_L1 processes, the range is from

46 Wilcoxon signed-ranks matched-pairs test: T=9; approximate p=0.75; r=0.14.
47 Welch's *t*-test on the ranks of the data: t[8.31]=0.97; p=0.36; 95 % CI [-3.32, 7.32].
48 Welch's *t*-test on the ranks of the data: t[10.00]=0.23; p=0.82; 95 % CI [-5.04, 6.04].

10 to 23 incidents per ten minutes and the median is 16 writing incidents per ten minutes. When the results are compared across tasks (Figure 9), writing frequency seems to be consistent for three of the six translators (Bidir2, Bidir4 and Bidir6). Bidir4 is the one with the highest frequency of writing in both tasks. An examination of the transcripts reveals that she produced a considerable number of what might be considered typing errors that per definition interrupted her production of new TT and increased the frequency of writing because she tended to immediately correct them (i. e. a revising incident) and then continue to write (i. e. a new writing incident). In terms of variation with regard to the two tasks, there is no significant difference between the processes of the bidirectional translators into L2 and into L1 with regard to writing frequency.[49]

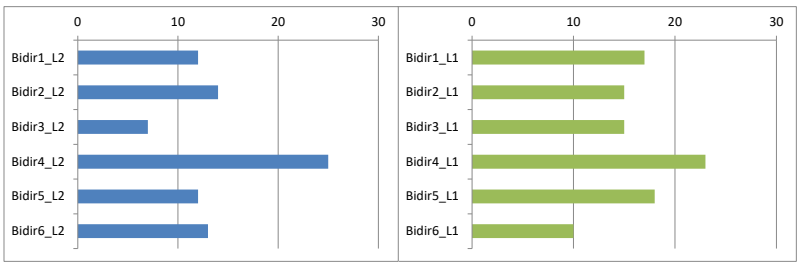


Figure 9: Frequency of writing in Bidir_L2 and Bidir_L1 processes ordered by translator (no./10 min)

For the UniEnglish_L1 processes, the frequency of writing ranges from 11 to 33 number of writing incidents per ten minutes, with a median of 23. A comparison of the writing frequency in the UniEnglish_L1 to that of the Bidir_L2 processes, which are associated with the same source text, shows no significant difference.[50]

For the UniGerman_L1 processes, the frequency of writing ranges from 15 to 29 writing incidents per ten minutes and the median is 22. A comparison of these results with those of the Bidir_L1 processes, which are based on the same ST, shows no significant difference.[51] All graphs are displayed in Appendix C.

In summary, the bidirectional translators tend to write with quite consistent frequency in both translation directions. When compared to the unidirectional translators with the same translation task, their writing frequencies seem to be comparable. As neither the number of writing

49 Wilcoxon signed-ranks matched-pairs test: T=5; approximate p=0.25; r=0.52.
50 Welch's *t*-test on the ranks of the data: t[9.31]=-1.98; p=0.078; 95 % CI [-8.42, 1.09].
51 Welch's *t*-test on the ranks of the data: t[9.95]=-1.74; p=0.11; 95 % CI [-8.25, 1.58].

activities nor their frequency varies much within or between the groups, the next logical question is whether the mean duration of writing incidents does.

5.3.3 Mean duration of writing incidents

In Figure 10, the mean duration of writing incidents for the Bidir_L2 and the Bidir_L1 processes are presented (see Appendix C for range and median for all task groups). It is noticeable that there is a tendency for the mean duration to be consistent for both translation directions for four of the six translators. For the two exceptions (Bidir5 and Bidir6), there seems to be a relation between mean duration and number of writing incidents (see Appendix C): in the process in which they performed more writing incidents, the mean duration of those incidents is lower. Overall, there is no significant difference between Bidir_L2 and Bidir_L1 with regard to the mean duration of their writing incidents.[52]

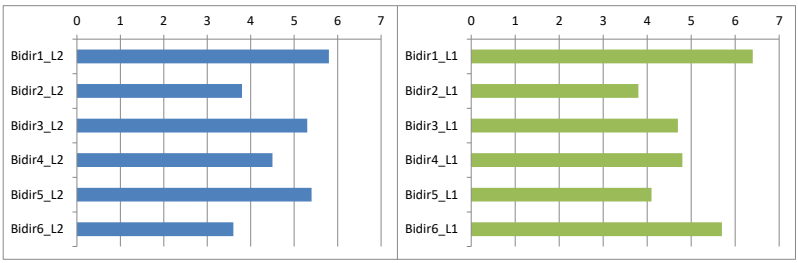


Figure 10: Mean duration of writing incidents (sec) for Bidir_L2 and Bidir_L1 processes, ordered by translator

When the results for the processes that are based on the same source texts are compared, the UniEnglish_L1 group and the Bidir_L2 group seem similar. The statistical analysis confirms that there is no significant difference in the mean duration of writing incidents between these two groups.[53] In the other intergroup comparison, the results for the UniGerman_L1 group also seem to correspond to those for the Bidir_L1. This impression is supported by a statistical test on the ranks of the data, which indicates no significant difference between the groups.[54]

In summary, the mean duration of writing incidents in the TP seems related to individuals' process patterns rather than to translation direction.

52 Wilcoxon signed-ranks matched pairs test: T=10; approximate p=0.92; r=0.05.
53 Welch's *t*-test on the ranks of the data: t[9.26]=-0.31; p=0.77; 95 % CI [-6.25, 4.92].
54 Welch's *t*-test on the ranks of the data: t[10.0]=0.23; p=0.82; 95 % CI [-5.04, 6.04].

In order to examine how much TT content is produced during these activities, their mean length in characters was calculated. The results are presented in the following section.

5.3.4 Mean length of writing incidents

For the mean length of writing incidents, I decided not to report the number of words but the number of characters because this allows for a comparison of the bidirectionals' individual behaviour, irrespective of the target language. Figure 11 presents the mean lengths of writing incidents in characters for the Bidir_L2 and the Bidir_L1 processes (for ranges and medians for all task groups see Appendix C). It is apparent that the mean length for both translation directions is remarkably consistent. As an example, in Bidir1_L2, a mean writing incident is 26 characters long and in Bidir1_L1, it is 24 characters long. Bidir6 is an exception, though: her mean number of writing incidents when translating into L2 is notably shorter than into L1. Overall, there is no significant difference between Bidir_L2 and Bidir_L1 with regard to the mean length of writing incidents.[55]

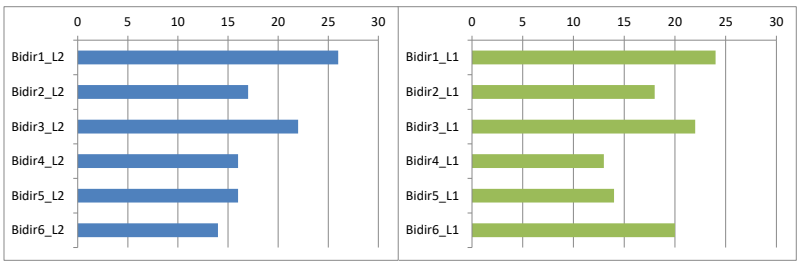


Figure 11: Mean length of writing incidents (char) for Bidir_L2 and Bidir_L1 processes, ordered by translator

When the results for the processes that are based on the same STs are compared, the UniEnglish_L1 group and the Bidir_L2 group seem similar. The statistical analysis confirms that there is no significant difference in the mean length of writing incidents between these two groups.[56] In the second intergroup comparison, the results for the UniGerman_L1 group are very similar to those of the Bidir_L1. The statistical analysis indicates no significant difference.[57]

55 Wilcoxon signed-ranks matched pairs test: T=9; approximate p=0.75; r=0.14.
56 Welch's *t*-test on the ranks of the data: t[10]=-0.96; p=0.36; 95 % CI [-7.37, 3.37].
57 Welch's *t*-test on the ranks of the data: t[9.93]=-0.46; p=0.65; 95 % CI [-6.55, 4.55].

In summary, the mean length of writing sequences in the TPs is considerably consistent across translation direction, that is to say for the bidirectional translators working into their L2 as well as working into their L1. The bidirectional translators also produced writing sequences of similar mean length as the unidirectional translators in both language versions. These results suggest that this measure is not related to directionality.

5.3.5 Summary of results for the writing measures

Overall, the bidirectional translators in this study tend to produce new TT in a manner that is comparable to that of the unidirectional translators. The analyses of the writing incidents suggest that the process descriptor writing may be more suitable to tracing individual behaviour irrespective of translation direction than to detecting differences that might be related to directionality.

5.4 Revising in the translation process

Another activity contributing to text production is editing or revising as it involves the reformulation of TT by deleting existing text and inserting new text. Generally, translators monitor their immediate output (i. e. what they are typing) very carefully throughout the process (e. g. Hvelplund 2017a, 59). This corresponds to what I observed in my study. Probably as a consequence of this monitoring, revision of the TT usually starts very soon after the text production begins and persists until the very end of the TP. In all 24 TPs in this study, the translators performed revisions throughout. It seems that the translators preferred to correct any obvious errors such as typos immediately after production rather than to revise them later on. At least one of the translators commented on this.[58] Immediate correction may help reduce cognitive load (i. e. the translator does not need to remember or remind herself later but can get rid of an easy-to-solve problem right away). While immediate correction reduces the number and length of writing incidents, it does not have an obvious influence on the number of revisions since errors would have to be corrected at some point anyway. The following section provides results of revision measures and comparisons between task groups. All comparisons are represented as graphs in Appendix C.

58 Bidir2 in her RVP on her L1 translation task: „Ja, Rechtschreibfehler, die werden sofort korrigiert." (My translation: Yes, spelling mistakes are corrected immediately.)

5.4.1 Total number of revisions

In the processes of the bidirectional translators working into their L2 (Bidir_L2), the number for *revising* ranges from 29 to 76 and the median is 73 (for the numbers for all 24 processes see Appendix C). The process with the lowest number of revisions seems to be an exception, as there is not much variation among the other group members. In the processes of the bidirectional translators into their L1 (Bidir_L1), there is a lot of variation between the group members: the range is from 37 to 112 with a median of 56. When the two translation directions are compared (Bidir_L2 to Bidir_L1), no particular pattern of individual behaviour seems to emerge. A statistical analysis shows no significant difference in the number of revisions between the translation directions.[59]

In the processes of the unidirectional translators working into English (UniEnglish_L1), the total number of revisions ranges from 31 to 89 and the median is 52. When the results for the UniEnglish_L1 group are compared to those for the Bidir_L2, which were elicited from the task involving the same ST, there is no significant difference between them.[60] In the processes of the unidirectional translators working into German (UniGerman_L1), the number of revisions ranges from 30 to 128 and the median is 48. Comparing the results for the UniGerman_L1 and the Bidir_L1 groups, which are based on the same source text, there is a tendency for the Bidir_L1 processes to have a higher number of revisions. However, the difference is not significant.[61]

In summary, there are no significant differences in total number of revisions during the TP within or between the groups. Therefore, this measure does not seem to be related to translation direction. The next section is concerned with the frequency of revising.

5.4.2 Frequency of revising

The combined measure of revision frequency comprises the number of revisions that a translator executes within ten minutes of process time. Within the Bidir_L2 processes, that frequency ranges from 11 to 32 revisions per ten minutes and the median is 19 revisions per ten minutes. Within the Bidir_L1 processes, the range is from 15 to 42 revisions per ten

59 Wilcoxon signed-rank matched-pairs test: T=10; approximate p=0.92; r=0.05.

60 Welch's *t*-test conducted on the ranks of the data: t[9.55]=0.32; p=0.76; 95 % CI [-4.71, 6.05].

61 Welch's *t*-test conducted on the ranks of the data: t[8.96]=0.67; p=0.51; 95 % CI [-4.10, 7.10].

minutes and the median is also 19 revisions per ten minutes. When the results for the two directions are compared (Figure 12), only the revision frequencies of Bidir3 and Bidir6 remain consistent across the two tasks. The frequencies for the other translators are much less consistent. Bidir4 has the highest revision frequency in both tasks. On the whole, there is no significant difference between the directions.[62]

Figure 12: Frequency of revising in Bidir_L2 and Bidir_L1 processes ordered by translator (no/10 min)

For the UniEnglish_L1 processes, frequency of revising ranges from 17 to 38 revisions per ten minutes and the median is 34. When the revision frequency of the UniEnglish_L1 group is compared to that of the Bidir_L2, there is a significant difference (Figure 13).[63] In other words, Bidir_L2 perform fewer revisions per ten minutes of task time than UniEnglish_L1. The fact that the former group also had longer process duration than but a similar amount of revisions to the latter has contributed to this difference. It seems that the unidirectional translators into English engage more in revising their TTs than the bidirectional translators who may engage more frequently in other process activities. An indication of differences in the proportion of revising with relation to the other activities between the two groups is already suggested in Section 5.2.2.

62 Wilcoxon signed-rank matched-pairs test: T=6; approximate p=0.35; r=0.43.

63 Welch's *t*-test conducted on the ranks of the data: t[9.58]=-2.25; p=0.049; 95 % CI [-8.57, 0.57].

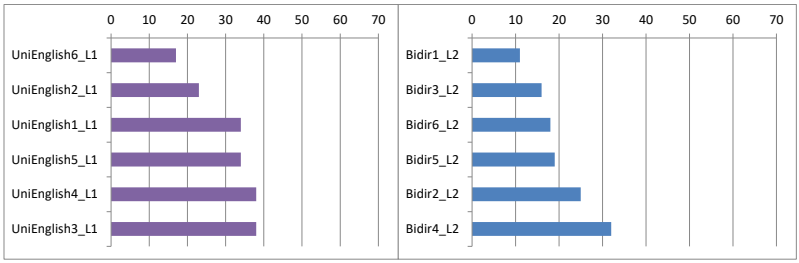


Figure 13: Frequency of revising in UniEnglish_L1 and Bidir_L2 processes in ascending order (no/10 min)

Within the UniGerman_L1 group, revision frequency ranges from 16 to 66 revisions per ten minutes and the median is 23. The UniGerman5_L1 process, with a revision frequency of 66, shows the highest frequency of all 24 processes by far. An examination of the transcript and other process measures reveals that the UniGerman5_L1 process is of average duration but contains a high number of revisions. The revisions do not, as might be expected, consist mainly of corrected typing errors but of a high amount of rephrasing. A comparison of the two groups UniGerman_L1 and Bidir_L1, whose processes are based on the same ST, shows no difference in revision frequency, as confirmed by the statistical analysis.[64]

In summary, the bidirectional translators translating into their L2 do revise less frequently than the unidirectional English translators do. This might be because they have fewer alternative solutions at their immediate disposal or rather tend to keep solutions that they have already produced.

5.4.3 Proportion of revisions per translation process phase

Translators revise text that they have already produced either when they are still writing their first version of the TT (referred to as the drafting phase) or when they are editing an existing version (revision phase). As they tend to work in cycles, it can well be the nth TT version they edit in an additional go. In my study, it was a short text that they had to produce. This is probably the reason I observed a linear style of working through the TT when revising it. With longer texts, it is likely that the translator first revises individual sections before revising the whole text in a linear fashion. The following section reports on the proportion of revisions performed in the drafting phase and in the revision phase. As per definition there are

64 Welch's *t*-test conducted on the ranks of the data: t[9.92]=-0.96; p=0.36; 95 % CI [-7.37, 3.37].

no revisions during the pre-writing phase, also known as the orientation phase, it is not included in the analysis.

Figure 14 shows the results for the bidirectional translators' processes into their L2 (Bidir_L2) and for their processes into their L1 (Bidir_L1). Into L2, two out of the six translators performed more than half of their revisions during the drafting phase as compared to during the revision phase. The other four translators performed between 36 % (Bidir5_L2) and 49 % (Bidir6_L2) of their revisions in the drafting phase. Into L1, all six bidirectional translators performed more than half of their revisions in the drafting phase.[65] Those two translators (Bidir1 and Bidir4) with the highest percentage of revisions in the drafting phase into L2 (83 % and 64 %, respectively) also had the highest percentage of revisions in the drafting phase into L1 (70 % and 82 %, respectively). For the other group members, the order is not consistent between tasks. When the two tasks are compared, in the Bidir_L2 processes, the proportion of revisions seems more balanced between the two phases than in Bidir_L1, where there is a slant towards the drafting phase. However, the difference is not significant.[66]

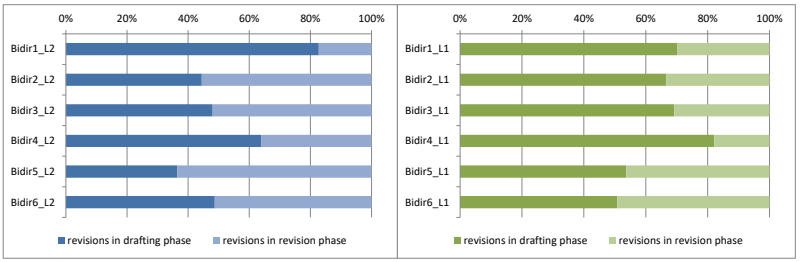


Figure 14: Proportion of revisions per TP phase (%) in Bidir_L2 and Bidir_L1 processes, ordered by translator

When the results for the Bidir_L2 processes are compared to those for the UniEnglish_L1[67] as they are based on the same task, the UniEnglish_L1 seem to do a higher proportion of their revisions in the drafting phase than the Bidir_L2 do. However, that difference is not significant.[68]

65 Range in Bidir_L2 for percentage of revisions in drafting phase: 36 %–83 %; median=48 %. Range in Bidir_L1: 51 %–82 %; median=68 %.

66 Wilcoxon signed-rank matched-pairs test on the proportion of revisions in the drafting phase: T=2; approximate p=0.075; r=0.81.

67 Range in UniEnglish_L1 for percentage of revisions in drafting phase: 36 %–96 %; median=76 %.

68 Welch's *t*-test on the ranks of the results for the proportion of revisions performed in the drafting phase: t[9.12]=1.53; p=0.16; 95 % CI [-2.06, 8.06].

In the other intergroup comparison, when the results for the Bidir_L1 processes are compared to those for the UniGerman_L1[69] as they are based on the same task, there does not seem to be a difference between the two groups: both share the tendency towards a higher proportion of revisions in the drafting phase than in the revision phase. In fact, there is no significant difference between the groups.[70]

In summary, when the translators work into their L1 (Bidir_L1, UniEnglish_L1 and UniGerman_L1), they tend to do a higher proportion of revisions in the drafting phase than in the revision phase. When they work into their L2 (Bidir_L2), the proportions seem to be more balanced between the two process phases. However, there are no significant differences. The following sub-section deals with a final measure of the process descriptor revising: the length of revisions.

5.4.4 Mean length of revisions in revision phase

In order to obtain more information on the amount of revision performed, one possibility is to calculate the length of the revision incidents in characters. I decided to apply this measure to revisions in the revision phase, as it is during that phase when revisions are in the translator's focus even though they tend to monitor their output throughout the whole TP (see also Fonseca 2015). All 24 TPs of this study included a revision phase where the translators performed deletions and insertions on their first TT versions in several revision cycles. In Figure 15, the mean length of those revisions for the Bidir_L2 and the Bidir_L1 processes are presented (see Appendix C for range and median of all groups). It is noticeable that in the majority of processes in both translation directions, the mean deletion length is comparable to the mean insertion length. This may be an indication that, on the whole, the translators only reformulate part of the text and do not add a lot of additional content to their first draft (i. e. most of the propositional content from the ST is already represented in the first draft). Although an insertion is not necessarily as long as the text sequence it replaces (i. e. the deletion), if the number and range of both revision types are comparable, this will result in a comparable mean length. Obvious exceptions are Bidir1_L2, Bidir3_L2, Bidir3_L1 and Bidir6_L1. These may be explained by taking a closer look at their deletions and insertions. As for Bidir1_L2, there was only one single deletion of 24 characters, which then

69 Range in UniGerman_L1 for percentage of revisions in drafting phase: 34 %–87 %; median=65 %.

70 Welch's *t*-test on the ranks of the results for the proportion of revisions performed in the drafting phase: t[9.66]=0.23; p=0.83; 95 % CI [-5.17, 6.17].

accounts for the high mean length. Bidir3 tended to work with variants, which she already introduced in the drafting phase. This means that at some point in the revision phase, she decided on a final version and deleted the alternatives, which accounts for the higher mean length of deletions. Into L1, she performed only a small number of revisions but reformulated large sections of text. In the Bidir6_L1 process, deleted expressions tended to be replaced by longer ones, which accounts for the mean insertion length exceeding the mean deletion length. As the mean length of both types of revisions well exceeds one character in both the Bidir_L2 and Bidir_L1 processes, it can generally be assumed that the bidirectional translators' attention during revision is not only on small units such as single characters but also on large ones such as clauses. An examination of the transcripts supports this assumption. A Wilcoxon signed-ranks matched-pairs test on the mean length of insertions as well as on the mean length on deletions in the revision phase indicates no significant difference between the two tasks with the bidirectional group of translators.[71]

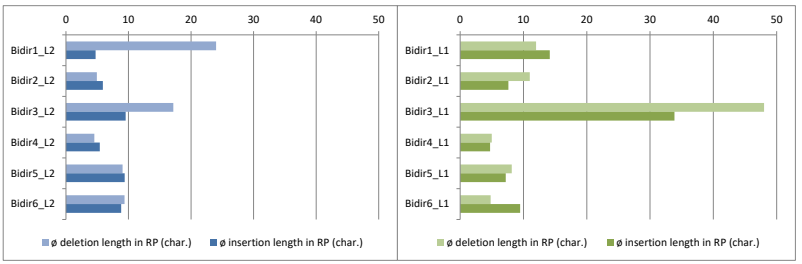


Figure 15: Mean length of deletions and insertions (in characters) in the revision phase of Bidir_L2 and Bidir_L1 processes, ordered by translator

Within each process of the UniEnglish_L1 group, the mean length for an insertion and the mean length for a deletion during the revision phase were strikingly similar (Appendix C). Compared to the results for the Bidir_L2 processes, there does not seem to be a difference between the groups. This is supported by a statistical analysis that compares mean insertion lengths and mean deletion lengths, respectively.[72]

71 Wilcoxon signed-ranks matched-pairs test in mean insertion length: T=6; approximate p=0.35; r=0.43. Wilcoxon signed-ranks matched-pairs test in mean deletion length: T=10; approximate p=0.92; r=0.05.

72 Welch's *t*-test on the ranks of mean insertion length: t[8.04]=0.62; p=0.55; 95 % CI [-4.17, 6.84].
Welch's *t*-test on the ranks of mean deletion length: t[9.60]=-0.62; p=0.55; 95 % CI [-6.84, 4.17]

Within each process of the UniGerman_L1 group, the mean length of an insertion and the mean length of a deletion during the revision phase were also very similar. UniGerman3_L1 was an exception with a much higher mean deletion length than mean insertion length. An examination of the transcript showed that the translator had simplified and heavily shortened her sentences in the revision phase, which resulted in the reported difference in mean lengths. When the results for the UniGerman_L1 are compared to those for the Bidir_L2 processes, there is no significant difference between the groups.[73]

Overall, there are no significant differences between the groups with the same STs or between the two tasks by the bidirectional translators, which may indicate that directionality is not an issue here. As seen above, mean revision length may rather be related to individual working style.

5.4.5 Summary of results for the revising measures

The total number of revisions does not seem to be related to translation direction. Of the six bidirectional translators, three performed more revisions into L2 and three into L1. Translation into German does not necessarily trigger a higher number of revisions than translation into English. Although in the German-English task, the bidirectional translators have a lower revision frequency than the unidirectional translators, both groups revise text sequences of similar average length. The translators' knowledge about and experience with text production seems to enable them to perform revisions as necessary.

The next section investigates the process descriptor *information seeking*. This descriptor refers to accessing and using online resources during the TP.

5.5 Information seeking in the translation process

Accessing and using online resources during the TP is an integral part of translation. In this study, when looking for information on the internet, both the bidirectional and the unidirectional translators generally took the search terms directly from the source text – independently of whether they looked for an expression in the TL or an explanation in the SL. They commented on this behaviour in the retrospective verbalisations. Moreover, they seemed

73 Welch's *t*-test on the ranks of mean insertion length: t[9.58]=-0.39; p=0.71; 95 % CI [-6.36, 4.70].
Welch's *t*-test on the ranks of mean deletion length: t[9.80]=-0.46; p=0.65; 95 % CI [-6.55, 4.55].

to prefer to type the search term into the search field rather than insert it by using the copy and paste functions. Usually, they did not write down an initial version before accessing online resources, although they often mentioned in their RVPs that they already had a solution in mind. Neither did they tend to produce variants after an information search – they usually decided on one solution only.

In the overview of the proportion that each type of TP activity comprises based on its number of occurrences, *information seeking* was responsible for the smallest share (see Figure 7). However, when it comes to the proportion in terms of time investment, information seeking covers 10 % to 48.2 % of total process duration (see Figure 8). The first measure analysed with respect to information seeking is the total number of occurrences in the 24 translation processes.

5.5.1 Total number of information searches

In the processes of the bidirectional translators working into their L2 (Bidir_L2), the number of information seeking activities ranges from 14 to 39 and the median is 23 (for the numbers for all 24 processes see Appendix C). There is some variation between the group members. In the processes of the bidirectional translators working into their L1 (Bidir_L1), the range is from 11 to 32 with a median of 18. When the processes are compared, there is a tendency for a higher number of searches in the TPs into L2 compared to those into L1. However, the difference is not significant.[74] When the processes of the same translators into both translation directions are compared, no apparent pattern of individual behaviour emerges.

In the processes of the unidirectional translators working into English (UniEnglish_L1), the total number of information searches ranges from 4 to 19 and the median is 12. A comparison of the UniEnglish_L1 processes and the Bidir_L2 processes, which are based on the same source text, shows that the former contain a significantly lower number of information searches.[75] The bidirectional translators seem either to have higher need for information searches for this particular task or to have integrated this general style of working into their translation process independent of direction.

In the processes of the unidirectional translators working into German (UniGerman_L1), the number of information searches ranges from 6 to 16 and the median is 8. Comparing the results for the UniGerman_L1 and the Bidir_L1 processes, which originate from the same task, the latter contain

74 Wilcoxon signed-ranks matched-pairs test: T=3; approximate p=0.12; r=0.71.

75 Welch's *t*-test on the ranks of the data: t[10]=2.90; p=0.016; 95 % CI [0.53, 8.80].

a higher number of information searches. However, the difference is not significant.[76]

In summary, the bidirectional translators performed significantly more information searches than the unidirectional translators in the German-English task did. This was not the case in the English-German task. In order to further examine the process descriptor information seeking, the frequency of the information searches during the TP is considered more closely.

5.5.2 Frequency of information seeking

As with the other process descriptors, the frequency with which information seeking is executed is reported in number of information searches per ten minutes. Within the Bidir_L2 group, the frequency of information seeking ranges from 5 to 9 searches per ten minutes and the median is 7 searches per ten minutes.[77] Within the Bidir_L1 processes, the range is from 4 to 9 searches per ten minutes and the median is 6 searches per ten minutes. The comparison of the processes for both translation directions shows that the information search frequency of four of the six translators is consistent for translation into L1 and into L2 (Figure 16). There is no significant difference between the translation directions.[78]

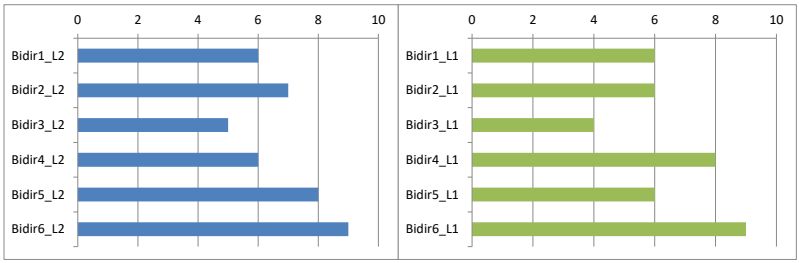


Figure 16: Frequency of information seeking in Bidir_L2 and Bidir_L1 processes (no/10 min), ordered by translator

A comparison of the information search frequency in the Bidir_L2 processes to that in the UniEnglish_L1 processes[79], which are associated

76 Welch's *t*-test on the ranks of the data: t[8.96]=2.02; p=0.074; 95 % CI [-1.09, 9.09].

77 I decided to round results to whole numbers here as translators only perform complete searches.

78 Wilcoxon signed-rank matched-pairs test: T=5; approximate p=0.25; r=0.52.

79 Range: 2–11 information searches/10 min; median=7 information searches/10 min.

with the same source text, shows greater variation within the UniEnglish_L1. Nevertheless, there is no significant difference between the groups.[80] In other words, the Bidir_L2 group and the UniEnglish_L1 group produced a similar number of information searches per ten minutes. A comparison of the results for the Bidir_L1 and the UniGerman_L1 processes[81], which are based on the same ST, shows a tendency to lower information seeking frequency for the latter. However, this difference is not significant either.[82]

In summary, the bidirectional translators perform a similar number of information searches per ten minutes to that of the unidirectional translators, irrespective of translation direction. In order to further examine the process descriptor information seeking, the mean duration of the information searches is considered as a measure.

5.5.3 Mean duration of information searches

Figure 17 presents the mean duration of information searches in seconds for the Bidir_L2 and the Bidir_L1 processes (for range and median for all groups see Appendix C). For Bidir4, Bidir5 and Bidir6, the mean duration of information searches remains remarkably stable across tasks. For the other three participants, it is considerably higher into L2 than into L1. The Bidir_L2 processes include significantly longer information searches than the Bidir_L1 processes.[83] Caution must be exercised when interpreting this difference between translation directions because the tasks involved different STs, whose topics might have required different amounts of research. A comparison with the unidirectional groups that translated the same source texts will show whether this difference persists.

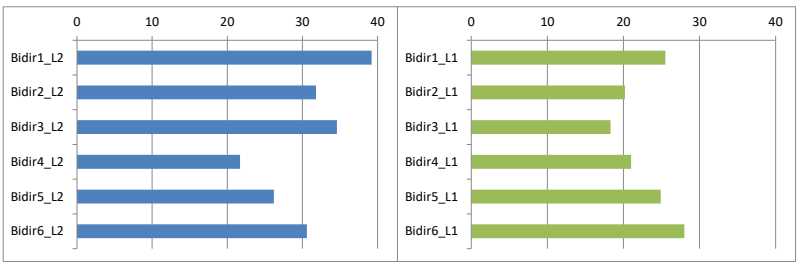


Figure 17: Mean duration of information searches (sec) for Bidir_L2 and Bidir_L1 processes, ordered by translator

80 Welch's *t*-test on the ranks of the data: t[8.15]=0.15; p=0.882; 95 % CI [-5.27, 5.94].
81 Range: 4–9 information searches/10 min; median=4 information searches/10 min.
82 Welch's *t*-test on the ranks of the data: t[6.56]=1.14; p=0.287; 95 % CI [-2.95, 7.62].
83 Wilcoxon signed-ranks matched-pairs test: T=not reported; p=0.028; r=1.0.

A comparison of the processes that are related to the same source text might suggest a difference between the UniEnglish_L1 group and the Bidir_L2 with regard to the duration of information searches as the means for the Bidir_L2 processes range from 22 to 39 seconds and those for the UniEnglish_L1 range between 12 and 22 seconds, except for UniEnglish2_L1[84] at 54 seconds. However, a statistical analysis indicates no significant difference between these two groups.[85]

In the other intergroup comparison, the results for the UniGerman_L1 group seem to more or less correspond to those for the Bidir_L1 (see Appendix C). This impression is supported by a statistical analysis on the ranks of the data, which does not indicate any significant difference between the groups with respect to mean duration of information search.[86]

In summary, the bidirectional translators perform significantly longer information searches when translating into their L2 than when translating into their L1. When compared to the other groups, the bidirectional translators tend to perform information searches of comparable duration. However, as in the group of the English unidirectional translators, one of the six results for mean duration could be considered an outlier, it remains unclear whether mean information search duration is related to directionality or not. So far, all measures on the process descriptor information seeking have been related to the total number of incidents. In the following sections, a selection of characteristics of these incidents is examined in more detail. To start with, the searches are scrutinised for potential subcomponents.

5.5.4 Ratio of single-step to multi-step information searches

When observing the translators and transcribing the activities in their TPs, I realized that the translators often perform multiple information searches in a row until they are satisfied with the result (i. e. until they find information that they can use to help them to continue producing the TT). During these series of information searches, they keep adjusting their search strategy, which involves both the type of online resource and the search term used. Below, the ratios between single- and multi-step information searches in the TPs of the four groups is presented and compared.

84 The UniEnglish_L2 process contains 4 information searches of 159 sec, 3 sec, 22 sec and 29 sec, respectively, which explains the high mean duration.

85 Welch's *t*-test on the ranks of the data: t[7.71]=1.98; p=0.084; 95 % CI [-1.09, 8.42].

86 Welch's *t*-test on the ranks of the data: t[6.88]=-0.15; p=0.883; 95 % CI [-5.94, 5.27].

In Figure 18, the results for Bidir_L2 and Bidir_L1 processes are presented.[87] Except for Bidir3_L1, multi-step searches account for the distinctly larger proportion of information searches within the TPs. For four of the six translators, the ratio remains remarkably stable across both translation tasks. However, when the processes are ranked according to their percentage of single-step searches, the order does not hold across translation direction. When analysed as repeated measures within a group, a statistical test does not reveal any significant difference.[88]

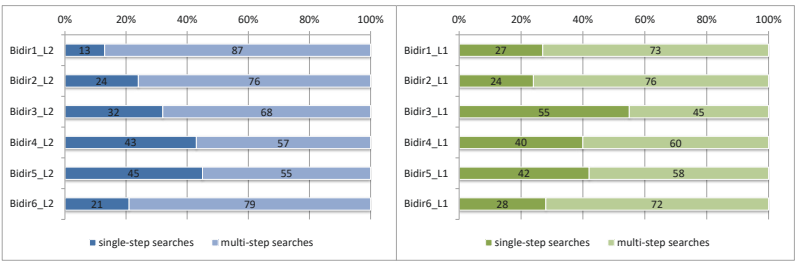


Figure 18: Ratio of single-step to multi-step information searches (%) in Bidir_L2 and Bidir_L1 processes ordered by translator

When the results for the UniEnglish_L1[89] processes are compared to those for the Bidir_L2, as they are based on the same source text, they are very similar with regard to range and median. In both groups, multi-step information searches dominate. A Welch's *t*-test on the ranks of the data confirms that there is no significant difference between these two groups.[90]

When the results for the UniGerman_L1[91] processes are compared to those for the Bidir_L1 (Figure 19), multi-step searches constitute the higher proportion in both groups. However, this imbalance is significantly more pronounced in the UniGerman_L1 than in the Bidir_L1 group.[92]

87 Range for single-step searches in Bidir_L2: 13 %–45 %; median=28 %. Range for single-step searches in Bidir_L1: 24 %–55 %; median=34 %.

88 Wilcoxon signed-ranks matched-pairs test on the percentage of single-step searches: T=6; approximate p=0.35; r=0.43.

89 Range for single-step searches in UniEnglish_L1: 20 %–50 %; median=27 %.

90 Welch's *t*-test on the ranks of percentages for single-step searches: t[10]=-0.4; p=0.698; 95 % CI [-6.20, 4.54].

91 Range for single-step searches in UniGerman_L1: 15 %–25 %; median=25 %.

92 Welch's *t*-test on the ranks of percentages for single-step searches: t[6.89]=4.02; p=0.006; 95 % CI [2.04, 9.29].

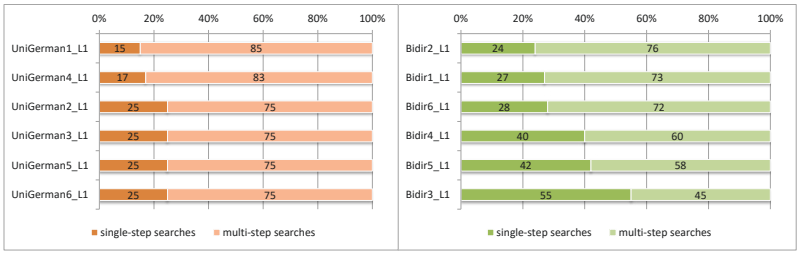


Figure 19: Ratio of single-step to multi-step information searches (%) in UniGerman_L1 and Bidir_L1 processes, in ascending order by single-step searches

In summary, both bidirectional and unidirectional translators tend to perform a higher proportion of multi-step searches than single-step searches during the TP. For the two groups of German-English translators, the proportion is similar. For the two groups of English-German translators, however, there is a significant difference: the bidirectional translators perform a higher percentage of single-step searches than the unidirectional translators do.

The next section reports on the variety of resource types that the translators accessed during their information searches. Again, intra- and intergroup comparisons are carried out.

5.5.5 Types of online resources used

Online resources that are accessed by the translator during the TP can be categorised into different types (see Section 4.3.6 in Chapter *Methodology*). During a multi-step search, the same type of resource can be accessed several times or in combination with other types. There are various reasons for accessing a specific type of resource. In their retrospective comments, the translators mentioned their motivation for using bilingual dictionaries, for example. According to the bidirectional translators, they accessed them during their L2 translation mainly to find TL equivalents and to confirm tentative solutions. In addition, they used them to research the search term's use in context, to see what was on offer and to help choose between variants (Hunziker Heeb 2015). In conclusion, a certain category of resource may serve a variety of needs that may not be attributed to its purpose on first sight. However, whether those needs can be served efficiently and adequately by any particular online resource is not the primary focus of this study.

In the following, all types of online resources that the translators accessed during their TPs are presented (for ranges and medians per group see Appendix C). The first comparison is of the bidirectional translators' use of resource types into L2 and into L1. Figure 20 shows the percentage of

each resource type used in an information search. At the top of the y-axis is the resource type with the largest occurrence (in average percentage) in the Bidir_L2 processes. The other resource types are then listed in roughly descending order. Whereas in the Bidir_L2 processes, ten different resource types appear, only eight appear in the Bidir_L1 processes. In both translation directions, bilingual dictionaries are the online resources that are accessed the most (40.2 % in Bidir_L2 and 43.1 % in Bidir_L1). These are followed by search engines for the Bidir_L2 (21.9 %) and parallel corpora (26.8 %) for the Bidir_L1. The intergroup comparisons that follow demonstrate whether the differences are related to the different STs or rather to directionality.

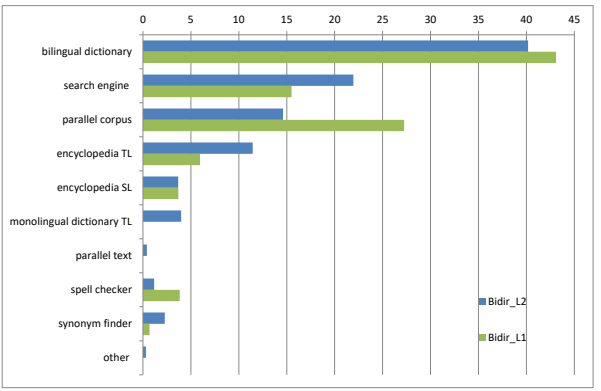


Figure 20: Mean use of types of online resources in Bidir_L2 and Bidir_L1 processes (%)

When the UniEnglish_L1 and the Bidir_L2 groups' use of online resources are compared, the similarity of results is remarkable (Figure 21). Both groups used bilingual dictionaries by far the most (mean of 40.2 % for Bidir_L2 and 36.3 % for UniEnglish_L1). These were followed by search engines, parallel corpora and encyclopaedias in the TL. Parallel texts such as news reports in the TL on the same topic were hardly ever used (less than a mean of 1 % for Bidir_L2 and never by UniEnglish_L1). However, it must be pointed out that other resource types can be used with the same purpose of finding topic-related information, formulations and terminology: parallel corpora and encyclopaedias in the SL and TL, namely Wikipedia, which often has entries on the same subject in German and in English. Interestingly, the largest difference in mean use is with the spell checker: whereas it was used in 1.2 % of the Bidir_L2's information searches, it was used in 6.4 % of the Unidir_L1's. On the whole, Bidir_L2 tended to use a greater variety of single online resources than the UniEnglish_L1, e. g. two different bilingual dictionaries instead of only one.

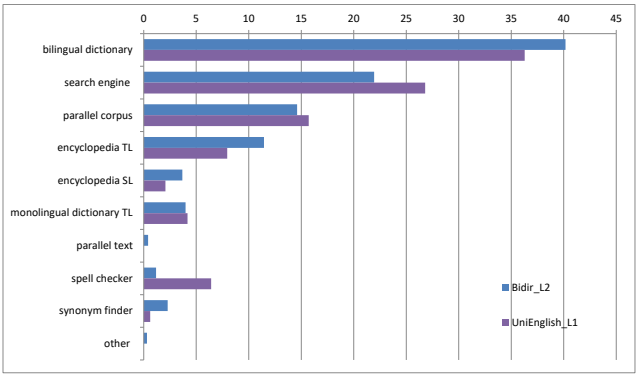


Figure 21: Mean use of types of online resources in Bidir_L2 and UniEnglish_L1 processes (%)

A comparison of the Bidir_L1 to the UniGerman_L1 group shows limited similarity with regard to resource use (Figure 22). Whereas in the Bidir_L1 processes bilingual dictionaries are the most-frequently used type of online resource (43.1 %), it is search engines (41.2 %) in the UniGerman_L1 processes. These are followed by parallel corpora for the Bidir_L1 (27.2 %) and bilingual dictionaries for the UniGerman_L1 (35.5 %), respectively.

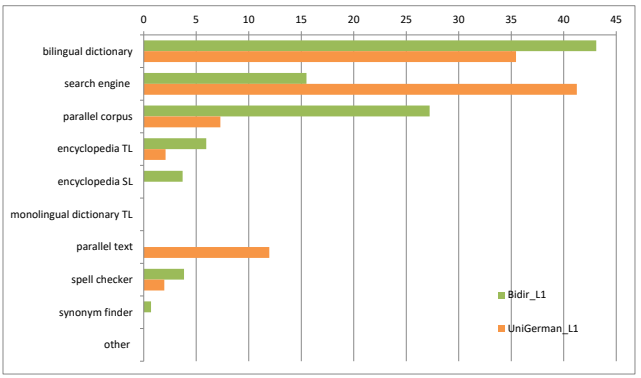


Figure 22: Mean use of types of online resources in Bidir_L1 and UniGerman_L1 processes (%)

In summary, the two groups with the German ST (Bidir_L2 and UniEnglish_L2) show rather noticeable similarity in terms of mean use of online resource types, which could point either to a need for a certain type of information or to a shared preference for certain resource types. By contrast, the two groups with the English ST (Bidir_L1 and UniGerman_L1) show a rather large difference in results. As both groups could have encountered similar

translation problems and therefore had similar needs for information searches and as they both translated into their L1, this difference could point to personal preference for certain resource types. It is also worth mentioning that monolingual dictionaries in the respective source language were not accessed during any of the processes. This could mean either that the translators did not need to address comprehension problems in the ST or that they addressed them by using another type of resource.

5.5.6 Summary of results for the information seeking measures

The bidirectional translators tend to do more information seeking in the L2 translation task than in the L1 task. The proportion of multi-steps to single-step searches is comparable in both tasks by the bidirectional translators. The unidirectional translators into English had a similar proportion whereas the unidirectionals into German had a much lower share of single-step searches than the other groups. Into their L2, the bidirectional translators use the greatest variety of resource types compared to the other direction and to the other groups. The bidirectionals' information searches were significantly longer into L2 than into L1. In comparison to the unidirectional translators' processes, no significant differences appeared, though.

It seems that the activity of information seeking is a rather entrenched behavioural pattern of the bidirectional translators, irrespective of directionality. It is possible that their L2 translation practice had an impact on this behaviour, although the scale of that effect remains to be determined. As the last of the four process descriptors, *pausing* will be investigated in the following sections.

5.6 Pausing in the translation process

For all process descriptors, the keystroke-logging data was used as the source. This means that *pausing* actually refers to gaps in the logs when no loggable activity is performed. Presumably, such pauses happen throughout the TP. The investigation of this descriptor starts with the calculation of its total number of occurrences. As in the sections on the other descriptors, the results are first presented for the two tasks the bidirectional translators did (i. e. Bidir_L2 and Bidir_L1) and then compared to those of the unidirectional translators into English (UniEnglish_L1) and into German (UniGerman_L1).

5.6.1 Total number of pauses

In the processes of the bidirectional translators working into their L2 (Bidir_L2), the number for all *pausing* incidents ranges from 20 to 48 and the median is 31 (for the numbers for all 24 processes see Appendix C). There is a lot of variation between the group members. In the processes of the bidirectional translators into their L1 (Bidir_L1), the range is from 20 to 40 with a median of 29. A comparison of the two directions shows that there is not much difference overall in range and median. This is supported by the statistical analysis.[93] When the processes of individual translators are compared, only two of them show consistency in their number of pauses for both translation directions.

In the processes of the unidirectional translators working into English (UniEnglish_L1), the total number of pauses ranges from 16 to 44 and the median is 25. A comparison of the UniEnglish_L1 processes and the Bidir_L2 processes, which are based on the same source text, shows a tendency for lower numbers of pauses in the former but the difference is not significant.[94]

In the processes of the unidirectional translators working into German, (UniGerman_L1), the number of pauses ranges from 20 to 40 and the median is 31. Comparing the results for the UniGerman_L1 and the Bidir_L1 groups, which are associated with the same source text, they are remarkably similar. This is supported by a statistical test that indicates no significant difference.[95]

In summary, the total number of pauses is only slightly higher for the bidirectional translators working into their L2 compared to working into their L1 or to the unidirectional groups, which are very comparable. Described next are the results for the frequency of pausing.

5.6.2 Frequency of pausing

The combined measure of pause frequency comprises the number of pauses that a translator has in ten minutes of process time. Within the Bidir_L2 processes, that frequency ranges from 8 to 15 pauses per ten minutes and the median is 10 pauses per ten minutes.[96] Within the Bidir_L1 processes, the range is from 7 to 15 pauses per ten minutes and the median is 13 pauses per ten minutes. When the results for the two directions are compared (Figure 23)

93 Wilcoxon signed-ranks matched-pairs test: T=3; approximate p=0.12; r=0.71.

94 Welch's *t*-test on the ranks of the data: t[9.22]=0.72; p=0.49; 95 % CI [-3.90, 6.90].

95 Welch's *t*-test conducted on the ranks of the data: t[9.66]=0.08; p=0.94; 95 % CI [-5.31, 5.64].

96 I decided to round results to whole numbers as translators do only perform complete pauses.

the pause frequencies of Bidir2, Bidir4 and Bidir6 are remarkably consistent across translation direction. The frequencies for the other translators are not consistent. Bidir6 has the lowest pause frequency in both directions. There is no significant difference between the processes in the two directions.[97]

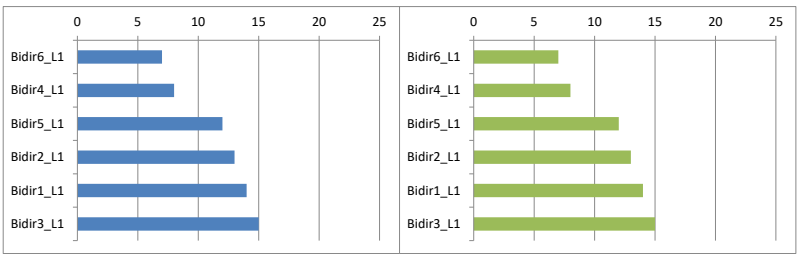


Figure 23: Frequency of pausing in Bidir_L2 and Bidir_L1 processes ordered by translator (no./10 min)

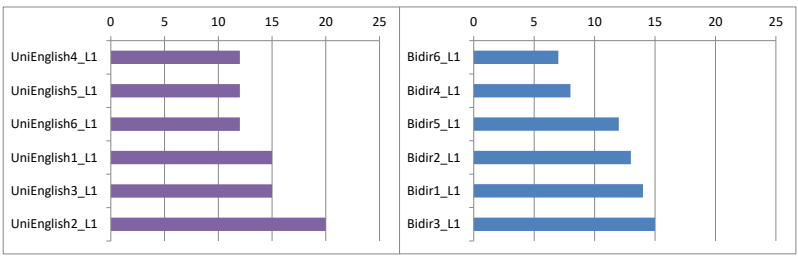


Figure 24: Frequency of pausing in UniEnglish_L1 and Bidir_L2 processes in ascending order (no./10 min)

When the pause frequency of the Bidir_L2 group is compared to that of the UniEnglish_L1[98], the observed frequency for Bidir_L2 is lower than for UniEnglish_L1 (Figure 24), and the difference is highly significant.[99] In other words, the bidirectional translators translating into their L2 (i. e. Bidir_L2) have less frequent long pauses than the unidirectional translators working into English do (i. e. UniEnglish_L1).

However, the other intergroup comparison (i. e. Bidir_L1 to UniGerman_L1[100]) showed no significant difference in pause frequency.[101]

97 Wilcoxon signed-rank matched-pairs test: T=4; approximate p=0.17; r=0.62.
98 Range: 12–20 pauses/10 min; median=14 pauses/10 min.
99 Welch's *t*-test conducted on the ranks of the data: t[10]=-5.56; p=0.0002; 95 % CI [-8.78, -3.22].
100 Range: 12–21 pauses/10 min; median=14 pauses/10 min.
101 Welch's *t*-test on the ranks of the data: t[9.60]=-1.53; p=0.16; 95 % CI [-8.06, 2.06].

In summary, the bidirectional translators have significantly fewer long pauses per unit of time when they translate into their L2 than the comparison group of unidirectional translators. Into their L1, the bidirectional translators' pause frequency does not differ significantly from that of the unidirectionals.

5.6.3 Mean duration of pauses

This study only considers pauses longer than five seconds, which has an impact on the measure of mean duration in seconds. In Figure 25, the results for the Bidir_L2 and the Bidir_L1 processes are presented (for range and median for all task groups see Appendix C). Two of the six translators had similar mean pause durations in both translation directions. Within each direction, there is some variation between the processes but overall, there is no significant difference between the Bidir_L2 and Bidir_L1 processes with regard to the mean pause duration.[102]

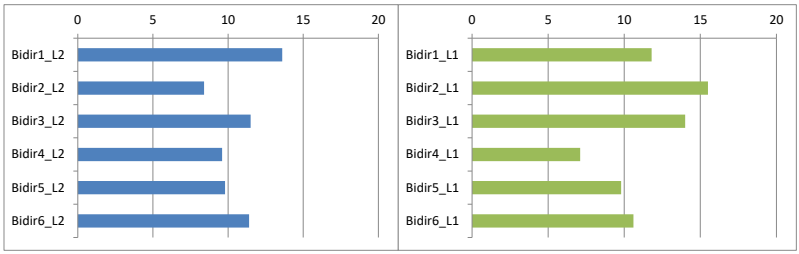


Figure 25: Mean duration of pauses (sec) for Bidir_L2 and Bidir_L1 processes, ordered by translator

In an intergroup comparison, the results for the UniEnglish_L1 and the Bidir_L2 processes, which are based on the same source text, do not seem to differ. The statistical analysis confirms that there is no significant difference in the mean pause duration between these two groups.[103]

Also in the other intergroup comparison, the general results for the UniGerman_L1 group seem to correspond to those for the Bidir_L1. This impression is supported by the statistical analysis of the ranks of the data, which indicates no significant difference between the groups.[104]

In summary, the mean duration of a long pause in the TPs is quite similar within and across the groups. For the majority of processes, the

102 Wilcoxon signed-ranks matched pairs test: T=10; approximate p=0.92; r=0.05.
103 Welch's *t*-test on the ranks of the data: t[10]=-0.46; p=0.65; 95 % CI [-6.55, 4.55].
104 Welch's *t*-test on the ranks of the data: t[9.97]=0.15; p=0.88; 95 % CI [-5.27, 5.94].

mean duration is longer than ten seconds. This means that long pauses are an integral part of professional translators' behaviour in this study.

5.6.4 Summary of results for the pausing measures

The statistical analysis of the different process indicators related to the pause activity only yields a significant difference for one case: the unidirectional English translators have much more frequent long pauses than the bidirectional translators working into their L2 have. This suggests that this measure might be related to translation direction. Within the bidirectional translation processes, no individual pause pattern seems to emerge.

5.7 Summary of results from the translators' processes

In total, 21 process indicators were measured and for 19 of them, statistical tests were run either for comparisons within the group of the bidirectional translators or between the groups who translated the same STs. Of the former, only one result was statistically significant, and of the latter, six results were significant. In the following, the results are summarised.

5.7.1 Intragroup comparisons

With respect to the intragroup comparisons between the bidirectional translators working into their L2 and the same translators working into their L1, the only significant difference is related to the mean duration of information searches: an information search in L2 translation takes significantly longer on average than an information search in L1 translation.

In revision behaviour, there is a non-significant tendency towards doing proportionately fewer revisions in the revision phase when translating into L1 than when translating into L2. In other words, when they translate into their L2, the bidirectional translators perform about half of their revisions in the drafting phase and half in the revision phase. When they translate into their L1, they perform more during the drafting phase and fewer during the revision phase.

The following measures showed remarkable consistency irrespective of translation direction for all or almost all of the bidirectional translators (i. e. all six or five of them): proportions of the four process activity types in terms of numbers; proportions of the four process activity types plus the category other activities in terms of duration; and the mean length of writing

incidents in terms of characters. The following measures showed consistency irrespective of translation direction for four of the translators: total number of writing incidents in the TP; the mean duration of these incidents; the mean length of an insertion in the revision phase; the frequency of information seeking; and the ratio between single- and multi-step searches.

5.7.2 Intergroup comparisons

Only one of the measures reported so far produced significant differences between the bidirectional and the unidirectional translators irrespective of language version: process duration. The bidirectional translators worked longer on their translations both into L2 and into L1 than the unidirectional translators did.

For the German into English version, three measures showed significant differences between the bidirectional translators into L2 and the unidirectional translators into L1: revision frequency; total number of information searches; and pause frequency. The bidirectional translators translating into their L2 revised less frequently than the unidirectional translators translating into their L1 did. The two groups did not differ with respect to number of revisions, though. The bidirectional translators translating into their L2 performed more information searches in online resources than the unidirectional translators translating into their L1 did. With respect to pause frequency, the bidirectional translators paused less often per ten minutes than the unidirectional translators did.

For the English into German version, only one measure resulted in a significant difference in the comparison between the bidirectional translators and the unidirectional translators: the relation of single-step to multi-step information searches. The bidirectional translators translating into their L1 performed a lower proportion of multi-step searches compared to single-step searches than the unidirectional translators did.

Table 11 lists all the measures reported on in this chapter in the order they were presented. For each measure, the results for the two intergroup comparisons are reported: When $p<0.05$, the difference between the two groups is considered significant. A cell with 'n. s.' indicates that the results were similar for both groups (i. e. that there was no significant difference between them). Striated cells indicate that the measure in question did not lend itself to statistical tests.

No.	Process descriptor	Bidir_L2 vs. Bidir_L1	Bidir_L2 vs. UniEnglish_ L1 (German-English)	Bidir_L1 vs. UniGerman_ L1 (English-German)
1	process duration	n. s.	p<0.05	p<0.05
2	character count	n. s.	n. s.	n. s.
3	total no. of process activities	n. s.	n. s.	n. s.
4	proportion of activity types	n. a.	n. a.	n. a.
6	total no. of writing incidents	n. s.	n. s.	n. s.
7	frequency of writing incidents	n. s.	n. s.	n. s.
8	mean duration of writing incidents	n. s.	n. s.	n. s.
9	mean length of writing incidents (char)	n. s.	n. s.	n. s.
10	total no. of revisions	n. s.	n. s.	n. s.
11	frequency of revising	n. s.	p<0.05	n. s.
12	proportions of revisions per TP phase	n. s.	n. s.	n. s.
13a	mean length of deletions in revision phase (char)	n. s.	n. s.	n. s.
13b	mean length of insertions in revision phase (char)	n. s.	n. s.	n. s.
14	total no. of information searches	n. s.	p<0.05	n. s.
15	frequency of information seeking	n. s.	n. s.	n. s.
16	mean duration of information searches	p<0.05	n. s.	n. s.
17	ratio of single-step to multi-step information searches	n. s.	n. s.	p<0.05
18	types of used online resources	n. a.	n. a.	n. a.
19	total no. of pauses	n. s.	n. s.	n. s.
20	frequency of pausing	n. s.	p<0.001	n. s.
21	mean duration of pauses	n. s.	n. s.	n. s.

Table 11: List of all the measures and the results for intergroup comparisons, in sequential order as reported in this chapter

In summary, most process indicators yield similar results for the bidirectional and the unidirectional translators who translated the same ST in the same language version. Therefore, it can be assumed that they are not related to translation direction. It is noteworthy that in the activities where the actual writing of the TT happens, the bidirectional and the unidirectional translators behave similarly. The indicators that yield significant differences between the bidirectional translators working into L2 and the unidirectional translators working into L1 (i. e. German-English) differ from those between the bidirectional translators and the unidirectional translators working into L1 (i. e. English-German) except for process duration. Hence, it is likely that a combination of several indicators is responsible for the difference in process duration for the bidirectional and the unidirectional translators, irrespective of language version and translation direction.

6 The translators' products

The purpose of a translation process (TP) is to create a translation product (i. e. the final target text or TT). In the professional setting, the focus is on the TT, since this is what the commissioner pays for and what the end-user reads. The TT is usually used and assessed detached from its source and its creator and therefore needs to make sense on its own. This chapter reports on the results of investigating the TTs produced by the bidirectional and unidirectional translators. It thereby addresses the research question 'In what ways are the products of professional L2 translation similar or different from the products of professional L1 translation?' The first section reports on general surface-level linguistic features of all 24 TTs, including global measures of the translation products. The subsequent sections then focus on product measures and quality evaluations of the twelve English TTs, as half of them are the products of L2 translation, the main topic of my study. As I did with the results of the main study, I also performed descriptive and inferential statistical analyses on the results of the substudy here.[105]

6.1 Global measures of the translation products

In order to describe the translation products that were produced in this study, global product indicators are introduced that allow for group comparisons. Two of the global product measures were: *number of TT characters* and *number of TT words*. These surface linguistic features inform about a general comparability of the translation products, which may be assumed but needs to be tested all the same. In a further step, the measures will be triangulated with process measures to define indicators of translation effort (see Chapter 7). The results of the global product measures are first reported for the TTs of the German-English translation task and then for the TTs of the English-German translation task. For a table with the actual numbers see Appendix D. The measure of *productivity* is used to compare how efficient the translators were in producing their target texts. The final global product measure reported in this section is the readability of the English TTs, which allows a direct comparison of L1 and L2 translations of the same source text.

105 Descriptive: range and median. Inferential: For intergroup comparisons, Welch's *t*-test on the ranks of the data is used as appropriate. The level of significance is 5 % if not stated otherwise. I performed two-tailed tests and report observed value of the t statistic (t), degree of freedom, p-value (p) and confidence interval (CI).

6.1.1 Number of target text characters

The number of characters in the English target texts of the bidirectional translators working into their L2 range from 663 (Bidir6_L2) to 763 (Bidir5_L2), with a median of 737. These numbers are similar to those of the English target texts by the unidirectional English translators into their L1, which range from 652 (UniEnglish1_L1) to 821 (UniEnglish2_L1), with a median of 727. The statistical analysis, which indicates no significant difference between the two groups, substantiates the comparability of results.[106]

The number of characters in the German target texts of the bidirectional translators working into their L1 range from 702 (Bidir2_L1) to 790 (Bidir5_L1), with a median of 738. These numbers are similar to those of the German target texts by the unidirectional German translators into their L1, which range from 685 (UniGerman4_L1) to 759 (UniGerman5_L1), with a median of 757. The statistical analysis, which indicates no significant difference, supports that the results are similar for the two groups.[107]

In summary, the bidirectional translators produced TTs that consisted of a similar number of characters to those produced by the unidirectional translators in the same language versions. Therefore, number of TT characters does not seem to be related to translation direction.

6.1.2 Number of target text words

The number of words in the English target texts of the bidirectional translators working into their L2 range from 105 to 133, with a median of 120. These numbers are similar to those of the English target texts by the unidirectional translators into their L1, which range from 113 to 136, also with a median of 120. The statistical analysis does not indicate any significant difference between the two groups.[108]

The number of words in the German target texts of the bidirectional translators working into their L1 range from 93 to 102, with a median of 96. These numbers are similar to those for the German target texts by the unidirectional translators into their L1, which range from 94 words to 99 words, with a median of 95. There is no statistically significant difference between the groups.[109]

In summary, the bidirectional translators produced TTs that consisted of a similar number of words to those produced by the unidirectional

106 Welch's *t*-test on the ranks of the data: t[8.99]=0.00; p=1.00; 95 % CI [-5.61, 5.61].
107 Welch's *t*-test on the ranks of the data: t[9.40]=0.15; p=0.88; 95 % CI [-5.27, 5.94].
108 Welch's *t*-test on the ranks of the data: t[9.99]=0.53; p=0.61; 95 % CI [-4.52, 6.85].
109 Welch's *t*-test on the ranks of the data: t[8.85]=0.42; p=0.68; 95 % CI [-4.24, 5.91].

translators in the same language versions. Therefore, the measure of number of TT words does not seem to be related to translation direction.

The analyses of the global measures of the translation products presented in this section suggest that the bidirectional and the unidirectional translators show very similar text production behaviour with regard to TT length, both in terms of number of characters and number of words.

6.1.3 Productivity

Product data is often combined with information about the process of text production in professional settings by reporting productivity, which measures how much TT the translator produces per unit of time. In other words, the temporal information of process duration is normalised by TT length, which increases the validity of the measure. Figure 26 presents productivity for the bidirectional translators working into their L2 and into their L1 in terms of words per minute. For the Bidir_L2 processes, productivity ranges from 2.5 to 5.3, with a median of 3.7. For the Bidir_L1 processes, it ranges from 2.7 to 4.9, with a median of 3.5. For four of the six bidirectional translators, productivity is rather similar in both translation directions. For the remaining two translators, productivity differs depending on direction: for Bidir3 it is lower into L2, whereas for Bidir4 it is much higher into L2. Both of those translators stated that they do most of their translation work into their L2 (as do two of the others). Overall, though, there is no significant difference between Bidir_L2 and Bidir_L1 with regard to productivity.[110]

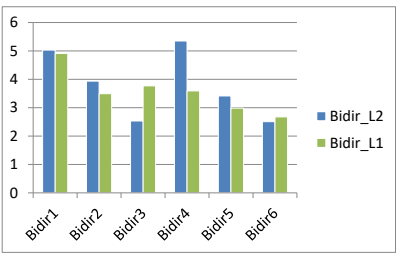


Figure 26: Productivity for the bidirectional translators for the translation processes into L2 and into L1, ordered by translator (words/min)

By contrast, the productivity of the UniEnglish_L1 group is significantly higher than that of the Bidir_L2 group, who translated the same ST (Figure 27).[111]

110 Wilcoxon signed-ranks matched pairs test: T=7; approximate p=0.46; r=0.33.
111 Welch's *t*-test on the ranks of the data: t[10]=3.32; p=0.008; 95 % CI [1.13, 8.87].

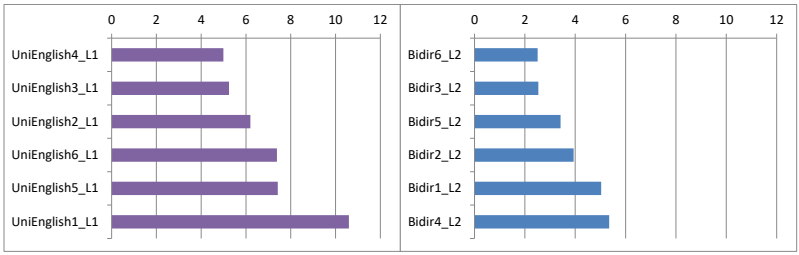


Figure 27: Productivity for UniEnglish_L1 and Bidir_L2 groups, in ascending order (words/min)

In addition, in the other intergroup comparison the UniGerman_L1 group shows a significantly higher productivity than the Bidir_L1 (Figure 28).[112]

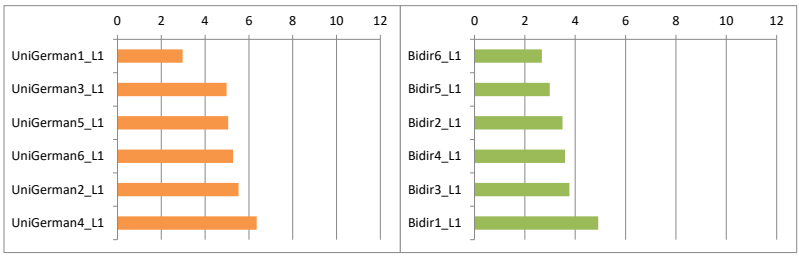


Figure 28: Productivity for UniGerman_L1 and Bidir_L1 groups, in ascending order (words/min)

In summary, the bidirectional translators produce a similar number of words per minute, irrespective of translation direction. Compared to the unidirectional translators, however, the bidirectional translators have a significantly lower productivity in both directions. This substantiates the process duration measures and differences between groups reported in Section 5.1.1, which were based on time information alone with no reference to the length of the final TTs.

6.2 Quality of the translators' products

The measure *product quality* was determined based on readability scores and on the results of a rating task carried out by 36 raters, who ranked the twelve English TTs according to perceived acceptability. In order to keep the task demands for the raters manageable, they only had to choose and

112 Welch's *t*-test on the ranks of the data: t[8.2]=2.55; p=0.03; 95 % CI [-0.04, 8.70].

rank the extreme positions (i.e. the three best from 1 to 3 and the three worst from 10 to 12, respectively).[113] I assigned a weighted score to each entry in one of those ranking positions. For each of the English TTs, I then calculated the sum of these scores to determine its level of acceptability. The product quality based on the evaluation by the total group of raters (n=36) is presented first. As the total rater group can be divided into two subgroups with respect to their first languages, the subgroups' results for product quality are then presented separately.[114]

6.2.1 Readability of the English target texts

In order to gain more information on the products of L2 translation and their potential differences from products of L1 translation, the readability of the English TTs was determined. Readability was operationalised with the Flesch Reading Ease formula. For details, see Section 4.3.6 in the *Methodology* Chapter.

Figure 29 presents the Flesch Reading Ease scores for the English TTs grouped by translation direction. A high score indicates easier readability than a low score.

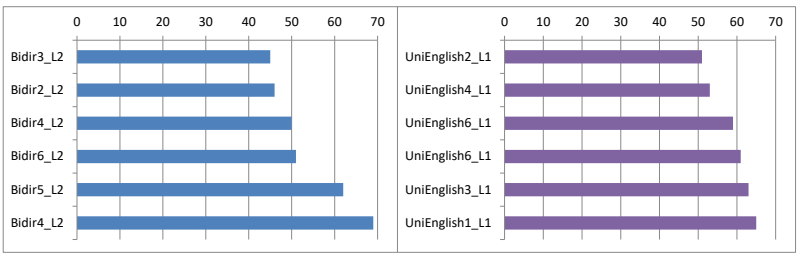


Figure 29: Flesch Reading Ease scores for English TTs of Bidir_L2 and UniEnglish_L1 groups, in ascending order

The scores for the TTs by the bidirectional translators into their L2 (Bidir_L2) range from 44.9 (Bidir3_L2) to 69.4 (Bidir1_L2), with a median of 50.3. The scores for the TTs by the unidirectional English translators into their L1 (UniEnglish_L1) range from 51.0 (UniEnglish2_L1) to 64.6 (UniEnglish1_L1), with a median of 59.8. A comparison of the two groups shows that the TTs produced by the Bidir_L2 group display a wider range of

113 All 12 English TTs appeared at least once in positions 1 to 3 as well as in positions 10 to 12, which points to either low interrater agreement or high similarity of the texts.

114 For a table with the acceptability scores per rater group see Appendix D.

readability scores than those produced by the UniEnglish_L1. On average, the TTs by the Bidir_L2 translators are more difficult to read than those by the UniEnglish_L1 translators. However, there is no significant difference between the two groups.[115]

When the TT reading ease scores are compared to those of a selection of US and UK newspapers (see Flaounas et al. 2013, 110), the TT with the highest score (69.4) is deemed easier to read than articles in *The Sun.* The TT with the lowest score (44.9) is comparable to articles from *The Wall Street Journal* with regard to readability. On average, the English TTs produced by the bidirectional translators are comparable to articles from *New York Post*, whereas those by the unidirectional translators, which are deemed easier to read, to articles from *Daily Mirror*. This means that all of the English TTs are easier to read than articles from *The Guardian*, which according to Flaounas et al. (2013, 110) have an average score between 30 and 35. This is noteworthy since, with respect to reading ease, *The Guardian* is considered comparable to the *Neue Zürcher Zeitung*, where the German ST was published. Both have a 'difficult' readability level, which is supposed to be suitable for college students in the US and students with completed A levels in the UK. The German ST has a reading ease score of 22 and is therefore deemed easy to understand only for readers at least the level of education of completed A levels.[116] This means that the English TTs produced by both groups of translators are considered easier to read – and implicitly to understand – than articles from a type of newspaper that the translation product was intended for.[117]

In summary, the bidirectional translators into L2 and the unidirectional English translators into L1 produced target texts of comparable readability. However, the target texts' reading ease scores were high compared to what could have been expected based on the instructions in the translation brief. Some of the participants (e. g. Bidir6_L2) used the website of *The Guardian* to get information on its style and on the use of certain phrases. The next section reports on the results of assessing the English target texts' quality.

115 Welch's *t*-test on ranks of the data: t[8]=1.32; p=0.22; 95 % CI [-2.51, 7.84].

116 For Flesch reading ease scores of German texts and newspapers see http://leichtlesbar.ch/html/.

117 The source text for the English-German translation task, which was an excerpt from an article originally published in *The Observer*, has a Flesch Reading Ease score of 49.3. In other words, it is also deemed 'difficult' but easy to read for US college students.

6.2.2 Acceptability of the English target texts

In the following, the acceptability scores for the twelve English TTs are presented (for operationalisation and data preparation see Section 4.3.6 in the *Methodology* Chapter). Figure 30 presents the results for all TTs in ascending order.[118] None of the TTs received either the maximum score (i. e. 432) or the minimum score (i. e. -432). Acceptability for the TTs produced by the Bidir_L2 ranges from -93 (Bidir1_L2) to 24 (Bidir5_L2), with a median of -36. Acceptability for the TTs produced by the UniEnglish_L2 ranges from -98 (UniEnglish2_L1) to 125 (UniEnglish3_L1), with a median of 53. Despite the difference in medians, the statistical analysis does not indicate a significant difference between the two groups.[119]

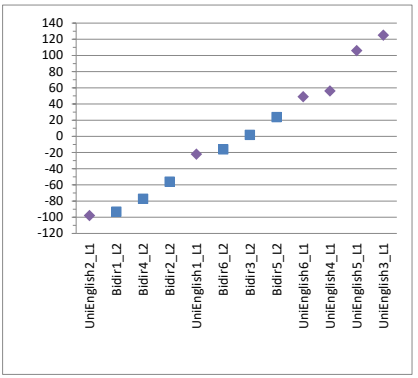


Figure 30: Acceptability scores for TTs of Bidir_L2 and UniEnglish_L1 in ascending order, based on evaluation by total rater group (n=36)

Overall, the bidirectional translators into L2 and the unidirectional translators into L1 produced TTs of similar acceptability. In other words, the relative quality of translation products of L2 translation processes is comparable to those of L1 translation processes. This suggests that product quality is not related to translation direction, at least for these professionals.

In the design of my substudy on the evaluation of product acceptability, I had two equal-sized rater subgroups in terms of their L1: While one of them had English as their L1 (L1 raters), the other half had English as their L2 or Lx (L2 raters).[120] As there have been assumptions expressed in the research literature that evaluators are influenced by whether the text they have to

118 As the scores for the bottom three positions were subtracted from those for the top three positions in the ranking, negative scores are possible.

119 Welch's *t*-test on ranks of the data: t[7.89]=1.53; p=0.17; 95 % CI [-2.06, 8.06].

120 For details on the raters see Section 4.3.2 in the *Methodology* Chapter.

assess is written in their L1 or not (see e.g. Hansen 2010, 394), the next logical step was to investigate whether the L1 raters and the L2 raters agreed in their assessment of product quality operationalised as acceptability.

Figure 31 presents acceptability for all twelve English TTs by the rater subgroups. The TTs are arranged in the same order as in Figure 30. The scores based on the L1 raters' evaluation (represented in Figure 31 by a circle) range from -84 to 58, with a median of 10. A comparison between the Bidir_L2 and the UniEnglish_L1 group suggests similarity. The statistical analysis supports that there is no significant difference.[121] Overall, the L1 raters did not distinguish between the texts produced by the bidirectional translators into their L2 and those by the unidirectional translators into their L1 as far as product acceptability is concerned.

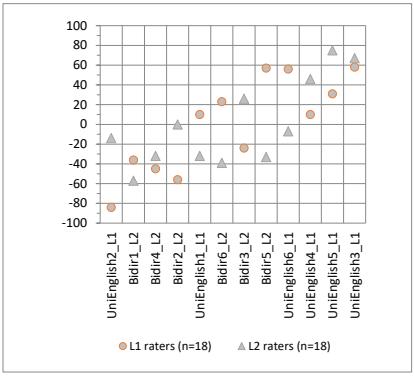


Figure 31: Acceptability scores for TTs of Bidir_L2 and UniEnglish_L1 based on evaluation by L1 raters and L2 raters, (in ascending order, based on evaluation by total rater group)

The scores based on the L2 raters' evaluations (represented in Figure 31 by a triangle) range from -57 to 75, with a median of -10.5. A comparison of product quality for the Bidir_L2 to the UniEnglish_L1 group suggests a tendency for lower scores for the former. However, the difference is not significant.[122] This means that the L2 raters did not seem to distinguish between the texts produced by the bidirectional translators into their L2

121 Range for target texts by Bidir2_L2: -56 to 57. Median=-30. Range for TTs by UniEnglish1_L1: -84 to 56. Median=21.
Welch's *t*-test on ranks of the data: t[9.89]=1.05; p=0.32; 95 % CI [-3.16, 7.49].

122 Range for target texts by Bidir2_L2: -57 to 26. Median=-33. Range for TTs by UniEnglish1_L1: -32 to 75. Median=20.
Welch's *t*-test on ranks of the data: t[9.85]=2.16; p=0.06; 95 % CI [-0.74, 8.40].

and those by the unidirectional translators into their L1 with respect to acceptability as an indicator of product quality either.

Overall, both rater subgroups failed to discriminate significantly between TTs produced in L2 translation processes and those produced in L1 translation processes with regard to quality. The subgroups showed comparable tolerance levels towards potential flaws in the TTs, irrespective of whether they had to assess them in their L1 or their L2.

A comparison of how every single TT scored per rater subgroup shows considerable intergroup variation: The difference in scores for the same TT ranges from only 9 (UniEnglish3_L1) to 90 (Bidir5_L2). This suggests that while there is high agreement with regard to acceptability for some of the TTs, there is little agreement for others. It remains unclear whether this range of disagreement for single TTs between the rater subgroups would be reduced with a higher number of raters or whether, on the contrary, it would be accentuated.

Overall, the potential target audience does not discriminate between products of L2 translation and products of L1 translation in terms of acceptability. Since the raters were asked to evaluate the TTs according to their own criteria, it would be of interest to know what those criteria were and whether they differ between the two rater subgroups. The following section reports on results that address these two questions.

6.2.3 Raters' evaluation criteria

In accordance with ecological validity, the raters were entirely free to set their own criteria for TT evaluation. They were simply informed that the texts had been produced by professional translators and were intended for publication in a quality English language newspaper. In the questionnaire that followed their ranking of the texts, they were asked about the evaluation criteria they had applied. They reported a total of 108 criteria (52 by the L1 raters and 56 by the L2 raters), which I then grouped into five categories. Table 12 presents the results for the total rater group as well as for the two subgroups of L1 raters and L2 raters in percentages based on the number of mentions.

evaluation category	*total rater group* (n=36)	subgroup *L1 raters* (n=18)	subgroup *L2 raters* (n=18)
readability and comprehensibility	24	29	20
grammar and syntax	24	23	25
idiomatic expressions and word choice	22	17	27
text composition and coherence	16	13	18
text type adequacy	14	17	11

Table 12: Self-reported evaluation criteria per rater group (%), in descending order by total rater group

When the evaluation criteria of all 36 raters (total rater group) are analysed, the two categories *readability and comprehensibility*, and *grammar and syntax* were applied most often (24 %). *Idiomatic expressions and word choice* cover 22 % of all criteria, *text composition and coherence* 16 %. Criteria subsumed under *text type adequacy* is used the least (14 %).

When the criteria for the two subgroups L1 raters and L2 raters are compared, the L1 raters applied readability and comprehensibility most often (29 %) while the L2 raters used idiomatic expressions and word choice most often (27 %). It seems that the L1 raters focus less on idiomatic language use than the L2 raters and more on overall quality of the texts. Here, the L1 raters' behaviour seems to be in accordance with their role as potential newspaper readers who probably take a rather holistic approach when judging the quality of an article that they are reading. By contrast, the L2 raters' tendency to use a more analytical or detailed (linguistic) analysis may reflect their approach to their own L2 text production evaluation or their own self-consciousness and experience as L2 learners of English. Both rater subgroups, however, regard correct grammar and syntax as the second most important criteria.

6.2.4 Correlation of acceptability and readability

In order to assess the relationship between the acceptability and the readability of the English TTs, a Kendall rank correlation test was performed (see Table 13). Results show that the correlations between readability and product quality were weak to very weak for all three aggregated ratings of product quality, although the L1 raters' evaluations were more related to readability than the L2 raters' evaluations were.[123] Nevertheless, the non-significance of these correlations suggests that the total rater group and the two subgroups L1 raters and L2 raters based their product evaluation

123 For the categorisation of correlation coefficients, see Chapter 8.

on more than just the texts' readability (as operationalised here) or else on other criteria altogether. This interpretation is in accordance with the fact that most of the raters actually listed several evaluation criteria.

	Product quality (by *total rater group*)	**Product quality (by *L1 raters*)**	**Product quality (by *L2 raters*)**
Readability	0.121	0.351	-0.137

Table 13: Rank correlations between readability and product quality of the English target texts (Kendall's τ)

6.2.5 Sequence of the translation tasks and product quality

Whereas the unidirectional translators only performed one translation task, the bidirectional translators performed two tasks. Therefore, it needs to be addressed whether the sequence of those two tasks could have affected the quality of the bidirectional translators' products. Since the source texts were on similar topics, those translators who performed the German-English translation as their second task may have profited from their prior exposure to the topic and corresponding activation of the topical vocabulary. Thus, they might have produced English TTs of higher quality than those translators who did the German-English translation first. If this were the case, the TTs of Bidir4_L2, Bidir5_L2 and Bidir6_L2 would have higher scores for product quality compared to the TTs of Bidir1_L2, Bidir2_L2 and Bidir3_L2 (for scores see Appendix D). To check this, I ranked the TTs within the Bidir_L2 group based on product quality, with the following result (in descending order): Bidir5_L2, Bidir3_L2, Bidir6_L2, Bidir2_L2, Bidir4_L2, Bidir1_L2. Since the translators who did the L2 translation after the L1 translation are not consistently at the top of the list, it can be concluded that there was no order effect with regard to product quality as operationalised in this study.

6.3 Summary of results from the translators' products

In terms of product quality as defined here, there was no significant difference between the English target texts produced by the bidirectional translators into their L2 and the English unidirectional translators into their L1. As indicators of product quality, readability and acceptability were measured. These results suggest that the potential target audience of the translations, which in this study was represented by readers of the

newspaper articles in English, did not differentiate between products of L2 translation and products of L1 translation in terms of acceptability.

Whether the raters actually followed their self-reported criteria when evaluating the target texts can only be assumed. This is in line with the widely admitted insight that translation product evaluation remains a subjective business (see for example Muñoz and Conde 2007). In her PhD dissertation, Nobs (2003) compared reader-raters' expectations towards translated texts and their actual evaluations of them. She found that although the raters claimed they had not used surface text features such as grammatical or orthographical errors as evaluation criteria for text quality, they actually had (Nobs 2003, 193).

7 The translators' effort

The target text is the product of the translation process and therefore, the two concepts translation process and translation product are closely related. The TT would not exist without the TP, which requires effort on the part of the translator to produce. After having analysed each of the items in isolation, this chapter focuses on the triangulation of process and product measures. It thereby addresses the third and last research question 'In what ways is the effort for translation of professional L2 translators similar to or different from the effort for translation of professional L1 translators?' To begin with, the chapter establishes and examines translator effort by means of four other indicators. Afterwards, it addresses relations between productivity, translator effort, professional experience and product quality. All results were subjected to descriptive and inferential statistical tests where feasible. Since the sample size is small and normal distribution of the data cannot be assumed, non-parametric statistical tests were used. For all types of comparisons, the statistical tests use the medians of the rank-transformed data. For intragroup comparisons, it is the Wilcoxon signed-rank matched-pairs test[124] and for intergroup comparisons, it is Welch's *t*-test on the ranks of the data.[125] The level of significance is 5 % unless stated otherwise. As in the other chapters on the study results, I first report on the results for the bidirectional translators into L2 and into L1 (Bidir_L2 and Bidir_L1, respectively) and illustrate them with a graph. Then, I compare the results for the Bidir_L2 and the unidirectional English translators into their L1 (UniEnglish_L1), as they translated the same source text. Afterwards, I report on the second intergroup comparison between the Bidir_L1 and the unidirectional German translators into their L1 (UniGerman_L1). Those intergroup comparisons are only illustrated with a graph if they yielded statistically significant differences. In Appendix E, all the comparisons are reproduced as illustrations for the sake of completeness.

Translator effort is operationalised as the triangulation of a selection of process and product data, which from the point of view of psychological reality are related to the effort a translator expends during the task. Triangulated data more adequately reflects the complexity of the translation task than any simple measure could do on its own. Putting a process measure in relation to a product measure is a way of normalising or standardising the measure. I determined five indicators of translator effort, which are considered to be indirect ones, on the grounds that a potential relation with

124 I report the test statistic (T), approximate p-value and effect size (r).

125 I performed two-tailed tests and report observed value of the t statistic (t), degree of freedom, p-value (p) and confidence interval (CI).

directionality would be justifiable and make sense. The aim of the analysis is to identify whether any of these indirect indicators is actually related to translation direction.

The five indicators of translator effort are: Character production effort, revision effort, information search effort, coordination effort and mean fixation duration. Character production effort suggests that producing more characters during the process than are actually present in the final TT is effortful. Revision effort implies that the amount of revision deemed necessary and therefore performed is an indicator of translator effort. I normalised the measure per ten TT words. Information search effort assumes that the extent of information seeking involved in TT production is related to effort. Coordination effort is based on the assumption that the TT needs to be synchronised somehow with the ST. This can be done during pauses, for example, which in this study are operationalised as being longer than five seconds. Mean fixation duration is the only indicator derived from eye-tracking data and is based on the premise that the eyes look at what the mind is concerned with, which makes sense in a cognitively demanding task such as translation. A difference in mean fixation duration in one group of translators compared to another would then point to a difference in effort. For all indicators, detailed definitions and operationalisations are presented in Section 4.4 of the *Methodology* Chapter. Details on data preparation and graphs of all comparisons are provided in Appendix E.

7.1 Character production effort

Character production effort indicates how many characters a translator produced during the process in order to create the final TT, which is an indirect indicator of translator effort. In other words, character production effort represents a process to product ratio in terms of characters. In this measure, I triangulate the global process measure *character count* and the global product measure *number of TT characters* (see Sections 4.2.6 and 4.3.6, respectively, for their operationalisation).

Within the Bidir_L2 group, the character production effort ranges from 1.5 to 2.8, with a median of 1.8. Within the Bidir_L1 processes, the range is 1.4 to 2.0, with a median of 1.6. When the results are compared between translation directions (Figure 32), character production effort is higher when working into L2 than into L1 for five of the six translators and consistent for one of them (Bidir4). Those translators who were among the top half with respect to character production effort into L2 also were among the top half into L1. Bidir6 has the highest measure in both translation directions: when working into L2, she had a ratio of 2.8 process characters per TT character and into L1, she had a ratio of 2.0 process characters per

TT character. On the whole, character production effort seems higher for L2 than for L1 translation. This is supported by the statistical analysis, which indicates a significant difference between the translation directions.[126]

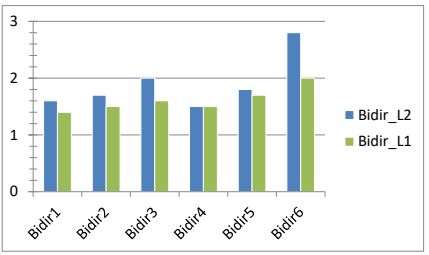


Figure 32: Character production effort for Bidir_L2 and Bidir_L1, ordered by translator (process char/TT char)

In contrast, character production effort is similar for the Bidir_L2 and the UniEnglish_L1[127] groups (see Appendix E), whose translations are based on the same source text. This is supported by the statistical analysis, which indicates no significant difference between the two groups.[128]

In addition, character production effort is also similar for the Bidir_L1 and the UniGerman_L1[129] groups, whose translations are based on the same ST (see Appendix E). This is supported by the statistical analysis, which shows no significant difference.[130]

In summary, the bidirectional translators exerted significantly greater character production effort when working into their L2 than when working into their L1. However, there is no significant difference when the bidirectional translators' character production effort is compared to that of the unidirectional translators. In other words, the bidirectional translators working into L2 produce a similar number of characters during their TP in relation to the characters in their TT as the unidirectional English translators. Therefore, it seems that this indicator of translator effort is not related to translation direction but more likely to characteristics of the ST for this translation version.

126 Wilcoxon signed-ranks matched-pairs test for Bidir_L2 and Bidir_L1: T=1; approximate p=0.046; r=0.91.

127 Range for character production effort: 1.3 to 2.0 process char/1 TT char; median=1.4 process char/1 TT char.

128 Welch's *t*-test on the ranks of the data: t[8.9]=1.98; p=0.079; 95 % CI [-1.09, 8.42].

129 Range for character production effort: 1.3 to 1.8 process char/1 TT char; median=1.5 process char/1 TT char.

130 Welch's *t*-test on the ranks of the data: t[9.53]=1.14; p=0.28; 95 % CI [-2.95, 7.62].

7.2 Revision effort

Reformulating and correcting existing text is cognitively demanding. On one hand, the microstructure needs to be correct and on the other, the changes need to be in accordance with the macrostructure chosen for the task at hand. The revision effort indicates the effort exerted to revise the text in order to arrive at the final TT. The assumption is that the translators have a similar view on what the TT should look like in terms of style, register and length. Revision effort is calculated as the number of revisions a translator performs per ten words of the final target text.

Within the Bidir_L2 group, the revision effort ranges from 2.2 to 7.2, with a median of 6.2. Within the Bidir_L1 group, the range for revision effort is from 4.0 to 11.8, with a median of 5.9. When the results are compared between translation directions (Figure 33), revision effort into L2 is lower than into L1 for three of the six translators and higher for the other three translators. Bidir4_L1 and Bidir5_L1 exerted by far the highest revision effort: They performed more than one revision per target text word. On the whole, revision effort is similar for both translation directions. A statistical test that compares the medians of the rank-transformed data indicates no significant difference.[131]

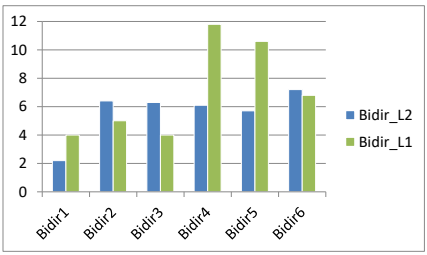


Figure 33: Revision effort for Bidir_L2 and Bidir_L1, ordered by translator (revisions/10 TT words)

A comparison of the revision effort for the Bidir_L2 to that for the UniEnglish_L1 group[132], which are associated with the same source text, shows greater variety for the latter. On the whole, however, their revision effort seems comparable (for graphs of all comparisons in this section see Appendix E). This is also suggested by the statistical analysis, which indicates no significant difference.[133] In addition, revision effort is also similar for the

131 Wilcoxon signed-ranks matched-pairs test: T=7; approximate p=0.46; r=0.33.
132 Range: 2.3–7.6 revisions/10 TT words; median=4.2 revisions/10 TT words.
133 Welch's *t*-test on the ranks of the data: t[9.26]=0.32; p=0.77; 95 % CI [-4.92, 6.25].

Bidir_L1 and the UniGerman_L1 groups[134], whose translations are based on the same ST. This is supported by the statistical analysis, which shows no significant difference between the groups.[135]

In summary, the revision effort of the bidirectional translators varies between translation directions. There does not seem to be stable individual performance on this measure. When compared to the unidirectional translators, revision effort seems comparable. These results point towards similar revision effort in both translation directions. The next section presents the effort involved in information searches in relation to TT words.

7.3 Information search effort

For the indicator of translator effort related to information searches, the process measure total number of information searches was used as a starting point. It comprises the total number of times that a translator performs an online information search on the same topic during the translation process. As discussed in Section 5.5.4, such a search can involve multiple accesses to resources or a single access only. As in this analysis I am interested in effort, I want to consider every single access to a resource. Therefore, I introduce the measure *number of queries*, which comprises the total number of times that a translator accesses an online resource and performs a search query. The numbers for all processes are listed in Appendix E. These results are then triangulated with the product measure *number of target text words*. The resulting indicator *information search effort* specifies how many times a translator accessed an online resource to perform an information search relative to a standardised measure of ten words of the final target text.

Within the Bidir_L2 group, the results for information search effort range from 1.8 to 8.8, with a median of 3.3. Within the Bidir_L1 processes, the range is from 1.3 to 7.1, with a median of 2.6. When the results are compared between translation directions (Figure 34), five of the six translators have a greater information search effort when translating into their L2 than when translating into their L1. Since for three of them (Bidir1, Bidir2 and Bidir3) L2 translation was actually the second task, the sequence of the tasks did not seem to have any facilitating effect on information search effort.

134 Range for revision effort: 3.1–13.1; median=5.0.

135 Welch's *t*-test on the ranks of the data: t[8.53]=0.79; p=0.45; 95 % CI [-3.78, 7.11].

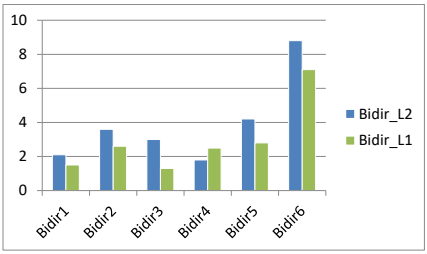


Figure 34: Information search effort for Bidir_L2 and Bidir_L1, ordered by translator (queries/10 TT words)

It is noteworthy that the rankings within the two translation directions are remarkably consistent: Those translators with high information search effort into L2 (compared to the other translators) tend to have high information search effort into L1 as well. As a consequence, it can be inferred that information search effort seems to be an individual trait. On the whole, information search effort is similar for Bidir_L2 and Bidir_L1, which is supported by the statistical analysis.[136]

By contrast, information search effort of the UniEnglish_L1 group (range: 0.7 to 2.5; median=1.7) is significantly lower than that of the Bidir_L2 group (Figure 35).[137] In other words, the bidirectional translators working into L2 perform a significantly higher number of search queries per target text word than the unidirectional English translators working into their L1.

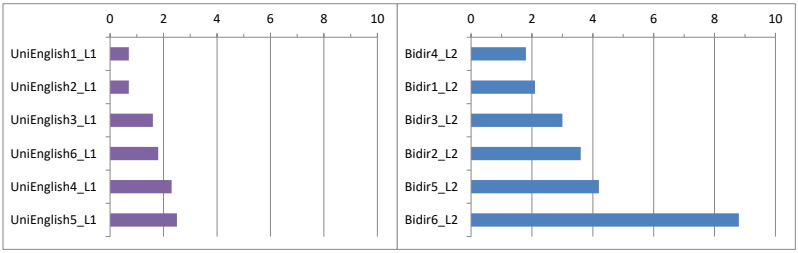


Figure 35: Information search effort for UniEnglish_L1 and Bidir_L2, in ascending order (queries/10 TT words)

A comparison of the results for the Bidir_L1 and the UniGerman_L1 groups, whose translations are based on the same ST, shows little variation among five of the six translators in both groups. One translator in each group, however, performs around three times more search queries per TT word than the others do (range: 0.9 to 6.1; median=1.2). On the whole, though,

136 Wilcoxon signed-ranks matched-pairs test: T=2; approximate p=0.075; r=0.81.
137 Welch's *t*-test on the ranks of the data: t[10]=2.90; p=0.016; 95 % CI [0.53, 8.80].

information search effort is similar for the Bidir_L1 and the UniGerman_L1 groups (i. e. no significant difference).[138]

In summary, the bidirectional translators have a similar information search effort in both translation directions. In other words, they perform a similar number of search queries per target text word. For the German-English translation processes, this difference is significant, which suggests that the information search effort indicator is related to translation direction. These results reflect what the process measure *total number of information searches* also showed (see *Main study results* Chapter): There is a difference between the bidirectional translators' information seeking behaviour when working into their L2 and the English unidirectional translators' when working into their L1. This may have to do in part with the amount of coordination the former group engages in between checking information results against the source and target text, as presented in the next section.

7.4 Coordination effort

While the other measures used as indirect indicators of translator effort so far were combined process-product measures, coordination effort is a triangulation of process measures: the process activity *pausing*[139] and the eye gaze data visualised in the screen recordings. During pauses, which in this study are operationalised as sequences longer than five seconds with no keyboard or mouse activity, translators may be looking at the screen, for example, because they are reading[140]. The item on screen where they focus their attention on during such pauses can therefore give an indication of translator effort.

Comparing the already produced TT to the ST is a common activity during the translation process. Translators seem to map the ST against the TT as well as the TT against the ST. This serves, for example, to ensure that no information has been unintentionally omitted in the TT or to reassure themselves about a passage in the ST while revising the TT. The indirect effort indicator coordination effort is operationalised as such switching between the ST and the TT during the process activity pausing. I determined the proportion of pauses to which this applies based on the visualisations of the

138 Welch's *t*-test on the ranks of the data: t[8.91]=1.98; p=0.079; 95% CI [-1.09, 8.42].

139 Range for *total number of pauses* in Bidir_L2: 20–48. Range in Bidir_L1: 20–40. Range in UniEnglish_L1: 16–44. Range in UniGerman_L1: 20–40. For a list of detailed numbers see Appendix C.

140 Here, reading can be understood as the eye gaze following characters from left to right in more or less sequential order.

gaze data overlaid on the screen recordings of the processes. Coordination effort is reported as a percentage of *total pause number*.

Figure 36 shows the coordination effort for the Bidir_L2 and Bidir_L1 processes.[141] For example, in the process of the bidirectional translator Bidir1 working into her second language (Bidir1_L2), the coordination effort is 61.5. This means that in 61.5 % of all pauses longer than five seconds the translator's eye gaze switches between the TT and the ST. This does not need to be in a systematic fashion or involve the same gaze time for both texts. Nevertheless, it can be inferred from the switches between the TT and the ST that the translator considers both texts and parts of them in her activity. During the Bidir1's process into her L1 (Bidir1_L1), she has exactly the same coordination effort of 61.5.

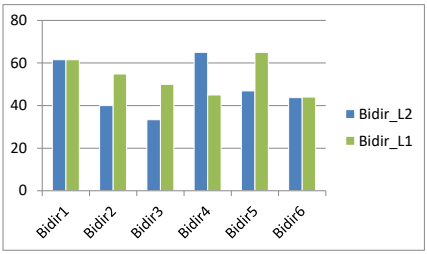


Figure 36: Coordination effort for Bidir_L2 and Bidir_L1 processes (%), ordered by translator

There is also remarkable consistency in coordination effort for another one of the translators (Bidir6) for both translation directions. Three of the other translators (Bidir2, Bidir3 and Bidir5) have a higher coordination effort into L1 while one translator has a higher one into L2. On the whole, there is no significant difference between Bidir_L2 and Bidir_L1 with regard to coordination effort.[142]

A comparison of the coordination effort in the UniEnglish_L1[143] to that in the Bidir_L2 processes, which are based on the same ST, shows similar results. The statistical analysis, which does not indicate a significant difference between the two groups, confirms this impression.[144]

A comparison of the coordination effort in the UniGerman_L1[145] to that in the Bidir_L1 processes shows a larger variation within the former but

141 Range for *coordination effort* in Bidir_L2: 33.3–65.0; median=45.3. Range in Bidir_L1: 44.0–65.0; median=52.4.

142 Wilcoxon signed-ranks matched-pairs test: T=7; approximate p=0.46; r=0.33.

143 Range for *coordination effort*: 28.6–66.7; median=50.0.

144 Welch's *t*-test on the ranks of the data: t[9.36]=0.08; p=0.94; 95 % CI [-5.48, 5.82].

145 Range for *coordination effort*: 25.0–73.9; median=48.8.

an overall similarity between the two groups. The statistical analysis of this comparison indicates no significant difference, either.[146]

In summary, the coordination effort, that is the synchronisation between ST and TT during pauses as defined in this study, is probably not related to directionality.

7.5 Mean fixation duration

The last of the five indicators of translator effort presented in this study is based on eye gaze behaviour during the whole translation process. A difference in mean fixation duration between tasks is assumed to indicate a difference in translator effort. In the following section, results are again compared between and across tasks.

Figure 37 shows mean fixation duration in milliseconds for the six bidirectional translators (Bidir1 to Bidir6) in both translation directions.[147] Two of the translators have longer mean fixation durations when translating into their L2 and four when translating into their L1. There is no relation between these results and the sequence in which the translators performed the two tasks. In general, mean fixation duration is remarkably consistent between translation directions. Bidir5 is an exception, with a difference of 153 milliseconds. Individual consistency is supported by the statistical analysis, which shows no significant difference with relation to translation direction.[148]

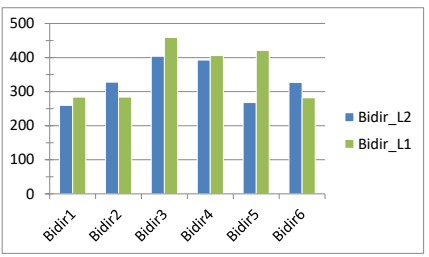


Figure 37: *Mean fixation duration* (ms) for Bidir_L2 and Bidir_L1 processes, ordered by translator

A comparison of mean fixation duration for the six bidirectional (Bidir_L2) and the six unidirectional translators (UniEnglish_L1[149]) in their German-

146 Welch's *t*-test on the ranks of the data: t[9.55]=-0.70; p=0.50; 95 % CI [-6.98, 3.98].

147 Range for Bidir_L2: 260–404 ms; median=328 ms. Range for Bidir_L1: 282–459 ms; median=345 ms.

148 Wilcoxon signed-ranks matched-pairs test: T=7; approximate p=0.46; r=0.33.

149 Range: 278–395 ms; median=335 ms.

English translation processes suggests a remarkable similarity between the groups. This is also suggested by statistical analysis, which shows no significant difference.[150]

The other intergroup comparison, the one between the six unidirectional translators (UniGerman_L1[151]) and the six bidirectional translators (Bidir_L1) for the English-German translation process showed more variation within the former group than the latter. Overall, though, there is no significant difference with regard to mean fixation duration.[152]

Overall, there is remarkable consistency in mean fixation duration for the bidirectional translators in both translation directions. Translator effort as indicated by mean fixation duration is also similar between the groups who translated the same ST. This suggests that translator effort in L2 translation is similar to translator effort in L1 translation with regard to mean fixation duration.

7.6 Summary of results from the translators' effort

Table 14 presents an overview of the results on translator effort. Of the five indirect indicators, only two show a significant difference for one of the three comparisons within and between groups: Character production effort is significantly lower for the bidirectionals working into their L1 than into their L2, and information search effort is significantly higher for the bidirectionals working into their L2 than for the unidirectionals working into their L1 in the German-English translation task. In conclusion, the majority of the results points towards a similar translator effort for L2 translation and L1 translation.

Indicator	**Bidir_L2 vs. Bidir_L1**	**Bidir_L2 vs. UniEnglish_L1**	**Bidir_L1 vs. UniGerman_L1**
Character production effort	$p<0.05$	n.s.	n.s.
Revision effort	n.s.	n.s.	n.s.
Information search effort	n.s.	$p<0.05$	n.s.
Coordination effort	n.s.	n.s.	n.s.
Mean fixation duration	n.s.	n.s.	n.s.

Table 14: Overview of comparisons of translator effort indicators (significance level indicated; otherwise non-significant)

150 Welch's *t*-test on the ranks of the data: t[9]=-0.15; p=0.88; 95 % CI [-5.94, 5.27].

151 Range: 168–661 ms; median=276 ms.

152 Welch's *t*-test on the ranks of the data: t[8]=0.54; p=0.60; 95 % CI [-4.40, 6.74].

The bidirectional translators expend significantly greater character production effort when working into their L2 than into their L1. The measure combines character production from writing, revising and information seeking activities, in other words, how many characters and blank spaces were actually typed during those activities. Task order did not seem to have had any effect on translator effort. This also means that the STs, which were about related topics, were still different enough so as not to cause a facilitating effect, which was the case in Ferreira's study of 2014. This again reinforces the reliability of the chosen measures as indicators of translator effort.

Revision effort varies for the bidirectional translators: While three expend more when working into L2, the other three expend more when working into their L1. As none of the analyses yields a significant difference, revision effort does not seem to be related to translation direction.

Information search effort is remarkably consistent for the bidirectional translators irrespective of translation direction. This could point to routine behaviour: Those translators who tend to invest a lot of effort in information seeking when working into their L2 also do so into their L1. One possibility is that they transfer their routines from their L2 translation activities to L1 translation. However, their information search effort is similar to that of the UniGerman_L1 group. Instead, it might be that the UniEnglish_L1 group is somehow special with respect to information search effort.

The indicator coordination effort is operationalised as the number of pauses during which the translator switches between gazing at the TT and gazing at the ST. Whereas the chosen threshold for a pause (i. e. more than five seconds) may be suitable to detect this type of pause activity for one translator, it may not be the case for another translator. As a consequence, if the coordination effort is low, this does not necessarily mean that the translator hardly switched between looking at the TT and the ST. It only means that she did not do it that often during these long pauses.

Mean fixation duration was calculated in this study over the whole translation task. Results for the bidirectional translators are remarkably consistent for both translation directions. Therefore, it seems that mean fixation duration is robust and indicative of an established processing routine that is independent of translation direction. An exception was Bidir5, who was the only participant who did not do both translation tasks in a row but performed the L1 translation task in the same project as the members of the comparison group (the unidirectional translators) did, that is 17 months before the L2 translation task. In the interview, she stated that she had not changed jobs between recordings and usually worked into her L2. With this in mind, one possible explanation for the large difference in mean fixation duration is that this measure varies from day to day or changes with translation experience. In his eye-tracking study on reading,

Rayner (1997) stresses that there is not only 'between reader variability' but also 'within reader variability' with regard to measures such as mean fixation duration. This means that a participant's fixation duration can vary considerably within a reading task. Nevertheless, as mean fixation duration is seen as an individual trait, a different approach might have increased the usefulness and validity of the measure for the intergroup comparisons. For example, the introduction of a base measure of mean fixation duration for each participant could be taken before the translation task from which the difference would then be calculated and used as an indicator of effort. Based on the results for mean fixation duration in this study at least, L2 translation and L1 translation seem to involve similar levels of translator effort.

8 Relations between characteristics of professional translation

This last chapter of results reports on relations between the previously triangulated process and product measures of translator effort, productivity and product quality as characteristics of professional translation. As presented in the overall study design in Figure 1, the additional variable *professional experience* is also used as a feature of professional translation. The level of professional experience is seen as an important factor in the relative proportion of routine tasks versus conscious tasks in a TP. As discussed in the literature review (see Section 3.4), it has been argued that performing routine tasks frees up cognitive resources that can then be devoted to complex problem solving. On the other hand, performing routine tasks could lead to higher risks of errors that go unnoticed by the translator. The relations between professional experience, translator effort and productivity are tested for all three groups (Bidir, UniEnglish_1 and UniGerman_L1). For the Bidir_L2 and the UniEnglish_L1 groups, the variable product quality[153] is also included in the analysis.[154] Product quality in this chapter refers to the acceptability scores reported in Section 6.2.

The analysis is performed by calculating and interpreting the correlation coefficient Kendall's τ. To facilitate the interpretation of the results, they are presented in matrices with the strength of correlation indicated as shown in Table 15. The exact correlation coefficients are listed in Appendix F.

Representation in results tables	**Degree of correlation**[155]	**Range of correlation coefficient Kendall's τ (unsigned)**
+/-	very weak negative or positive correlation	0.0-0.19
++/--	weak correlation	0.2-0.39
+++/---	moderate correlation	0.40-0.59
++++/----	strong correlation	0.60-0.79
+++++/-----	very strong correlation	0.80-1.0

Table 15: Representation of correlation coefficient ranges

153 Product quality was determined for the target texts of the German-English translation processes only.

154 The operationalisations of the components are reported in the Sections 4.2.6 and 4.3.6 and the results in the corresponding results sections in Chapters 6 and 7.

155 This interpretation is based on http://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf. According to Mellinger and Hanson (2016, 191), Spearman's correlation test is comparable to that of Kendall's. I considered reporting the interpretation rather than the coefficients in the text more reader-friendly.

Table 16 lists the correlations between the five translator effort measures, productivity and professional experience for the bidirectional translators when working into their L2 (Bidir_L2) and into their L1 (Bidir_L1). Whenever possible, information is presented only once in the matrix (redundant cells are indicated with n. a.).

	Measure	Productivity		Professional experience	
		Bidir_L2	Bidir_L1	Bidir_L2	Bidir_L1
Translation effort	Character production effort	- - - - -	- - - -	- - - -	- - - -
	Revision effort	- - -	- - -	- - - -	- -
	Information integration effort	- - - -	- - - -	- - - - -	- - -
	Coordination effort	+ + + +	+ +	+ +	+
	Mean fixation duration	-	+ +	+	+
Productivity		n.a.	n.a.	+ + + +	+ + + + +

Table 16: Correlations between *translator effort*, *productivity* and *professional experience* for the bidirectional translators

As reported in Section 7.1, *character production effort* is significantly higher in Bidir_L2 than in Bidir_L1 processes. This is also reflected in its correlation with productivity, which is very strong for Bidir_L2 and slightly less so for Bidir_L1. For both translation directions (Bidir_L2 and Bidir_L1), three of the five effort indicators correlate negatively with productivity: The more effort the translator exerts for character production, revision and information integration, the less productive she is. This makes sense, as these indirect effort indicators involve process activities, which in turn take time to perform and therefore decrease productivity. For both translation directions (Bidir_L2 and Bidir_L1), the same three effort indicators also correlate negatively with professional experience: The more experienced a translator is, the less effort she expends on character production, revision and information search.

The effort indicator *coordination effort* correlates positively with productivity and professional experience. A high proportion of pauses in which the translator's gaze switches between TT and ST is related to high productivity and long experience as a professional, with the correlation with the productivity indicator particularly strong for translation into L2.

The effort indicator *mean fixation duration* correlates only very weakly to weakly with productivity or professional experience. This may suggest that productivity and professional experience have little bearing on the assumed concentration or effort as measured by mean fixation duration.

As for the relation between productivity and professional experience, there is a strong positive correlation in both translation directions. These results support the assumption that the more experienced translators are, the higher their productivity.

The same correlations and comparisons were also carried out for the data from the processes of the bidirectional translators into their L2 (Bidir_L2) and the unidirectional English translators into their L1 (UniEnglish_L1). Table 17 lists the correlations between the five translator effort indicators, productivity and professional experience for both groups. In addition, it lists the correlations between those measures and product quality, which had been determined for their TTs in the substudy (see Section 6.2.1).

	Measure	Productivity		Professional experience		Product quality	
		Bidir_ L2	UniEnglish_ L1	Bidir_ L2	UniEnglish_ L1	Bidir_ L2	UniEnglish_ L1
Translation effort	Character production effort	- - - - -	-	- - - -	-	+ + +	+ + + +
	Revision effort	- - -	- - - -	- - - -	- -	+ +	+ +
	Information integration effort	- - - -	- -	- - - - -	- - -	+ + +	+ + +
	Coordination effort	+ + + +	+	+ +	+ +	- -	+ +
	Mean fixation duration	-	+ +	+	- -	+	+ + +
Productivity		n.a.	n.a.	+ + + +	-	- - -	- -
Professional experience		+ + + +	-	n.a.	n.a.	- - - -	- -

Table 17: Correlations between translator effort, productivity, professional experience and product quality for the Bidir_L2 and the UniEnglish_L1 groups

The three effort indicators related to character production, revision and information search correlate negatively with both productivity and professional experience for both groups. However, they correlate positively with product quality. While coordination effort correlates very positively with productivity for the Bidir_L2, it shows only weak correlations in the rest of the comparisons. The reasons for this are not obvious. As for mean fixation duration, it has stronger correlations with the other indicators in the UniEnglish_L1 group than in the Bidir_L2. Again, there is no straightforward interpretation. It is noticeable that while productivity shows a strong positive correlation with professional experience for the Bidir_L2 group (as might be expected) it correlates negatively and only weakly for the UniEnglish_L1. Possibly, since productivity was significantly higher for the UniEnglish_L1 anyway (see Section 6.1.3), the level of professional experience may not have had a great impact. As the UniEnglish_L1 are all staff translators in the same large LSP, high productivity is probably even a job requirement.

As regards the relation between productivity and product quality, they correlate negatively, albeit it only weakly to moderately, for both groups. This suggests that high productivity does not necessarily lead to high

product quality. Contrary to expectations, professional experience also correlates negatively with product quality for both groups: strongly for the Bidir_L2 (τ=-0.60) and weakly for the UniEnglish_L1 (τ=-0.20). If there actually is a causal relation between the two indicators, it may well be that it is overly simplistic to infer that longer professional experience necessarily leads to higher product quality. Other factors such as the volume translated (i.e. part-time vs. full-time workload) certainly have an influence on professional experience. Alternatively, in line with Göpferich's (2013) attempt to explain similar results in her longitudinal study by applying dynamic systems theory, translators may experience a stagnation of their translation competence and therefore also of their level of performance after a certain number of years as a professional.

The final set of comparisons in this section concern those for the bidirectional translators working into their L1 (Bidir_L1) and the unidirectional German translators working into their L1 (UniGerman_L1). Table 18 lists the correlations between the translator effort indicators, productivity and professional experience for both groups. As translation into the L1 was not the focus of this study, the German target texts were not assessed, and thus there are no measures for product quality.

	Measure	Productivity		Professional experience	
		Bidir_L1	UniGerman_L1	Bidir_L1	UniGerman_L1
Translation effort	Character production effort	- - - -	- - -	- - - -	- -
	Revision effort	- - -	- -	- -	- -
	Information integration effort	- - - -	- - -	- - -	- -
	Coordination effort	+ +	+ +	+	+ +
	Mean fixation duration	+ +	- - -	+	- -
Productivity		n.a.	n.a.	+ + + + +	+ + + +

Table 18: Correlations between translator effort, productivity and professional experience for the Bidir_L1 and the UniGerman_L1 groups

Overall, it is noticeable that the strengths of the correlations are quite similar between the two groups of translators. This is in contrast to the previous comparison, where there were more differences between the two groups. What the two comparisons have in common is that the directions of correlations largely coincide. The three effort indicators character production effort, revision effort and information search effort correlate negatively with productivity and professional experience for both the Bidir_L1 and the UniGerman_L1. This makes sense, as they are all related to text production. The correlation directions for the last effort indicator, mean fixation duration, do not coincide: While they are positive in relation to

productivity and professional experience for the Bidir_L1, they are negative for the UniGerman_L1. However, none of them is particularly strong. As in the other comparisons, interpreting the results for mean fixation duration does not seem to be straightforward.

Three indicators of translator effort correlate negatively with productivity in all four conditions (Bidir_L2, Bidir_L1, UniEnglish_L1 and UniGerman_L1), one indicator correlates positively and one varies. This suggests that the different types of effort have a different impact on productivity.

For the bidirectional translators, there are stronger correlations between professional experience and the effort indicators than for the unidirectional translators. For the bidirectional translators, there are also stronger correlations between productivity and the effort indicators than for the unidirectional translators.

9 Synthesis and outlook

This chapter summarises the main findings of the study by answering the research questions and reports on relations between directionality and characteristics of professional translation . It then reflects on these findings by discussing and embedding them in Cognitive Translation Studies. It also acknowledges the study's limitations, establishes its relevance, suggests further research paths and widens the scope to suggest reconceptualisations and potential implications beyond the academic discipline.

9.1 Answers to my research questions

Taking into consideration the premises made, the definitions used and the participants chosen in this study, I present the answers to the three research questions posed in Chapter 2 as follows. Differences are reported first, as they are the ones that could be confirmed by statistical measures.

The first research question concerns directionality and the translation process, specifically what the similarities and differences are in the processes of professional translators who are accustomed to working into their L2 and L1 compared to those exclusively working into their L1. Differences are reported first, as they could be substantiated by statistical measures. Consequently, if the statistical analysis did not yield a significant difference, results are reported as potential similarities between the translation directions. In this study, the translation task has been conceptualised as a complex situated cognitive activity that involves many types of actions and whose observation yields a large volume of data. Its investigation has therefore included a large number of process descriptors. In total, 21 process descriptors were measured (for detailed results see Chapter 5).

With respect to the two translation directions, the bidirectional translators' processes only differ significantly in the mean duration of the performed information searches (Table 11, descriptor no. 16). When the bidirectional translators worked into their L2, they spent more time on average in online information searches than when they translated into their L1. Since their behaviour in that respect was not different from the comparison groups, the unidirectional translators, for the respective language version, it can be assumed that mean duration of information searches is probably related to the task at hand rather than to directionality. In other words, the German source text may have prompted longer searches than the English one. Translating the German source text might have presented more challenges that the translators tried to resolve by performing longer information searches compared to those presented by the English source text. This result

is in contrast to findings reported by the PACTE group (for a discussion of their study see Section 3.2). The professional translators in their study did not perform significantly longer searches into their L2 than into their L1 with regard to PACTE's five predetermined potential translation problems (Kuznik 2017b, 228). However, not all of the translators in the PACTE study were accustomed to translating into their L2.

When the bidirectional's processes into their L2 are compared to those of the unidirectional English translators who performed the same German-English translation, they differed significantly in four descriptors (Table 11, no. 1, 11, 14, 20): process duration, frequency of revising, total number of information searches and frequency of pausing. Process duration is the most prominent difference as it also holds for the English-German translation. The bidirectional translators took significantly longer to finish both their German-English and their English-German translations than the unidirectional translators did (Table 11, no. 1). While the significantly higher number of information searches may potentially account for the difference in the German-English task, there is no single factor that emerges as potentially responsible for the difference in the English-German task. It is more likely attributable to the combination of a higher number and a larger duration of various process activities, which led then to longer process durations of the bidirectional translators. However, since that difference was not found *within* the group of bidirectionals in connection with translation direction, directionality probably has little or no effect on process duration. This result corroborates findings that professional bidirectional translators do not spend significantly more time on L2 translation than on L1 translation tasks (e. g. da Silva et al. 2017, 128 and Ferreira 2012, 86; for a discussion of their studies see Section 3.2). It therefore seems reasonable to assume that the difference in time on task (i. e. process duration) is related to the fact that one group of translators works in both translation directions whereas the other groups work in one direction only. In light of this result, characteristics of the bidirectional and the unidirectional groups in this study are revisited below (for their description as participants see Section 4.2.2).

The other three differences in the translation processes of the bidirectionals working into their L2 and those of the English unidirectional translators are the frequency of revisions and of pauses as well as the total number of information searches. In the German-English translation task, the L2 translators revise and pause less frequently but perform more information searches than the L1 translators do. Buchweitz and Alves' (2006) finding that L2 translation involves a higher number of long pauses starting at five seconds thus cannot be corroborated by my results. In any case, these differences do not seem to be related to directionality alone, as they do not emerge as significant when the two tasks of the bidirectional translators are compared. The high number of information searches that the

bidirectional translators perform into their L2 has an effect on the frequency with which they perform revisions and pauses. It seems that information seeking is more important to the bidirectionals in that task than it is to the unidirectional translators.

The process descriptors related to revision seem to depend on working style: If a translator tends to produce an already elaborate target text version initially, this may result in shorter revisions during the revision phase. Alternatively, if she works with variants, these have to be decided upon and hence the surplus variant deleted at some point during the TP. Translators also vary with respect to whether they tend to delete and rewrite whole sections or prefer to use the cut and paste functions in their editors instead. The relation between working style and mean insertion or revision length is complex, since a high number of revisions may result in a low mean length and vice versa. Moreover, process descriptors related to revision may also depend on the language in which they are performed. It can be assumed that revisions in English texts comprise fewer characters on average since average word length tends to be shorter than in German texts. The number of revisions performed, however, may be higher in German than in English as more morphological adjustments in agreement, case, etc. might need to be made in German if a revision is made elsewhere (e. g. changing a German verb may require a shift from the accusative to the dative, which would then mean morphological changes in the articles, adjectives and noun of the grammatical object). This type of detailed linguistic analysis might be interesting to do in a follow-up study. The fact that the mean length for deletions and insertions is larger than one character in all but one process (i. e. UniEnglish4_L1[156]) suggests that the translators revised their target text versions beyond the correction of spelling mistakes, irrespective of translation direction. This impression was confirmed in a close examination of the transcripts.

While information searches consisting of multiple steps dominate those consisting of a single step in all groups for both tasks, its ratio in the unidirectional German translators' processes is significantly higher than in the bidirectional translators' processes into L1. It is difficult to draw inferences on the reasons for this difference without actually examining the information searches in detail (e. g. their linguistic content from the logging data and their supposed goals from the retrospective commentaries). The decision to perform several searches in a row most likely depends on the result of the search that has just been performed. The reasons for unsatisfactory results are probably manifold. For example, the first choice of which resource to use or the search term entered first may have produced a result that was

156 UniEnglish4 only performed two deletions and two insertions in the revision phase, each consisting of one character.

deemed inadequate. By adjusting her choice of resource or her search term, the translator then increases her chances of finding useful information.

In my study, information seeking accounts for more than 25 % of the translators' time.[157] This result is roughly in line with what has been found in previous studies. For example, Hvelplund (2017a, 76) reported that information seeking accounted for nearly 20 % of professional translators' total task time. In her study on the use of web resources, Gough (2015, 129) found that her participants spent an average of 33 % of their task time on information searches. Of course, this amount is task-dependent and also whether the translator is familiar with the topic of the source text.

The preference for a particular type of online resource may depend on the types of translation problems that have to be solved, such as culturemes (Olalla-Soler 2018) or titles (Ehrensberger-Dow and Massey 2013). In his study on five culturemes in a ST that professional translators had to translate into their L1, Olalla-Soler (2018, 1306) found that the three most frequently used resources were general search engines (37.8 %), bilingual dictionaries (22.4 %) and encyclopaedias (18.8 %). Gough (2015, 141) was concerned about the fact that the professional translators who participated in her survey did not seem to know all the resources they were asked about even though most of these resources covered multiple languages and therefore seemed useful for translation. As a consequence, translators may not necessarily use the most efficient resources for certain types of translation problems.

However, the type of resource used is not necessarily an indication of the type of translation problem that triggered its use. To categorise the translation problems that each translator sought information about or the type of queries that each translator performed, every query would have to be analysed in detail. Gough (2015, 177–178) equates entering a search term in the source language with source text-orientation and a translation approach focussed on formal equivalence whereas entering a search term in the target language is target-oriented focussing on functional equivalence. It seems problematic to come to this conclusion based solely on the language of the search term. My participants' reasons for doing research, which they provided during retrospection, seem to point in that direction. While the majority of their search terms were directly taken from the source texts, many participants claimed that they already had had a solution in mind and only wanted to check it or look for variants. Based on their comments, about 40 % of the bidirectionals' information searches in the L2 translation task may be attributed to looking for an equivalent and about 25 % to confirming a tentative solution (Hunziker Heeb 2015). Looking

157 Bidir_L2: mean=24 %; UniEnglish_L1: mean=19 %; Bidir_L1: mean=27 %; UniGerman_L1: mean=27 %.

for inspiration also seems to be a common reason for consulting bilingual dictionaries (Enríquez 2014, 24–25).

I agree with Gough, however, that the choice of resource can be attributed to habit and/or preference (Gough 2015, 178). It is probably also faster and cognitively less demanding to just enter the source text term in a search field. Possibly the translator does not want to restrict herself too much at that point in case something more appropriate turns up. Alternatively, she may simply need a trigger to generate a solution. Risku (1998, 175) states that bilingual dictionaries and thesauri can only be used to trigger what the translator already knows by association. Otherwise, these resources are not reliable enough to produce good solutions as they do not provide any information about use in context or in a specific communicative situation. Kirsh (2009, 301) mentions self-cueing as a strategy for problem-solving, which people have been reported to use when playing Scrabble.

The total number of process activities that the translators perform during the TP is also similar within and between the groups. This suggests that the number of process activities is not related to directionality either. In terms of the investigated process activities *writing*, *revising*, *information seeking* and *pausing*, the following similarities were found: all four types of activities are present in all TPs with the number of information seeking activities generally accounting for the lowest proportion and revising for the highest proportion. In general, the actual production of the target texts in terms of lengths and mean durations of writing and revising as well as the proportion of revisions performed in each phase are comparable, irrespective of translation direction. This comparability also relates to the number and duration of pauses as well as the duration of information searches.

Overall, the results of the process analyses that were carried out to answer the first research question suggest that there are by far more similarities than differences between the translation processes of bidirectional translators compared to those of unidirectional translators, at least for the professional translators who participated in this study (for detailed results see Section 5.7). Furthermore, none of the process descriptors measured in this study was convincingly found to be related to directionality.

As the analysis of the translation processes in this study showed very few differences attributable to directionality, the second research question addresses whether this also applies to the translation products. A comparison of the target texts per task shows that they are similar with regard to number of characters and words (details see Section 6.1), irrespective of translation direction.

When productivity is considered, it seems that although only few differences in process behaviour between the bidirectional and the unidirectional translators were significant, they added up to significantly lower productivity of the bidirectional translators (see Section 6.1.3).

However, there was no difference in productivity between the bidirectional translators translating into their L2 and into their L1. To the contrary, their productivity was quite stable in both translation directions, suggesting that directionality *per se* does not seem to affect productivity. The difference in productivity between the bidirectional and unidirectional groups rather seems to be related to the fact that the bidirectional group is used to translating both into and out of their L1 whereas the other group only translates into their L1.

In order to investigate the assumption that translating into one's L2 results in poorer products than when translating into one's L1, the quality of the English target texts was evaluated. As a first measure, the readability of the target texts was compared. Results show that the average readability of the target texts produced by the bidirectional translators into their L2 was similar to that of the texts produced by the English unidirectional translators into their L1 (see Section 6.2.1). However, both groups of translators produced texts that are considered easier to read than articles from the type of publication the translations were actually intended for. It seems that the translators had set priorities other than matching readability when considering the needs of the target audience.

The quality of the target texts produced by both groups of translators was also similar when evaluated by potential users. The user-raters failed to make a distinction between texts produced in L2 translation and those produced in L1 translation. This finding contributes to answering the question by Schwieter and Ferreira (2017, 77) of whether "translations into English produced by English native speakers really [are] of a superior quality compared to those produced by nonnative English speakers". The answer seems to be a straightforward 'no, they are not'. McAlester (1992, 292–293) assumed that a translator who is an L2 speaker of English may be in an as good as or even a better position than an L1 speaker when it comes to translating for an international audience with English as their L2, that is for an ELF situation. The results of my study show that the bidirectional translators working into their L2 were able to meet the target text users' requirements as well as the L1 translators were. My findings also corroborate Pokorn's (2005, 117) results, which suggest that users with the target language as their L1 were not able to recognise whether a text was produced by a translator working into her L1 or into her L2. She concludes that "the assumption that every native speaker is able to rapidly detect any non-member of his/her linguistic community, when confronted only with a written document, has no solid foundation" (Pokorn 2005, 117). Interestingly, the users in my study who had the target language as their L2 rated the target texts by the bidirectional translators into their L2 slightly lower than those by the English unidirectional translators, albeit not significantly (see Section 6.2.2). As L2 speakers of English, they might have been overly sensitive to what they perceived as non-standard structures in

the English texts, for example from their own experience as language learners (see also Murray 2003 or Derwing, Rossiter and Ehrensberger-Dow 2002).

Since I did not evaluate the quality of the target texts that the bidirectional translators produced into their L1, the question of whether product quality is linked to directionality cannot be answered with certainty. Potentially all of the evaluated English texts could have been of low or even unacceptable quality as only their relative quality was assessed. However, this seems highly unlikely as they were produced by trained professionals who translate for a living, some as staff translators. Moreover, none of the raters commented on the potential unacceptability of any of the target texts.

In general, the results suggest there are no significant differences in the target texts produced by the professional translators who participated in this study that could be attributed to directionality. This was contrary to the expectations based on the review of some of the literature and especially the views explicitly or implicitly expressed by some professional associations and language service providers.

The third research question addressed whether there are differences in the amount of effort that translators have to invest, depending on directionality. In light of the answers to the first and second research question, this can be more precisely expressed as whether the bidirectional translators had to exert more effort to achieve the same level of quality when working into their L2 and their L1. In this study, the cognitive effort that translators expend in performing a translation task has been referred to as translator effort and investigated by analysing the five indirect indicators character production effort, revision effort, information search effort, coordination effort and mean fixation duration (Chapter 7). For the majority of indicators for translator effort, no significant differences were found between the task groups (see Table 14 in Section 7.6).

Within the group of bidirectional translators, character production effort was the only indicator that showed a significant difference in the two translation directions. This effort indicator represents the general effort invested in text production in relation to the length of the target text as it is based on the number of characters produced during the whole task. Therefore, the result that the bidirectional translators expended significantly more character production effort in their L2 translation than in their L1 translation task (see Section 7.1) suggests that when working into their L2 they were more engaged in producing text than when working into their L1. However, as they did not expend more of this type of effort compared to the English unidirectional translators, character production effort does not seem to be related to directionality. It is probably more related to the German-English task, which seems to have required a higher amount of effort of this type. Whether this is related to any difference in source text difficulty between the two translation tasks is unclear.

In the intergroup comparisons, information search effort is the sole indicator that resulted in a significant difference. In the German-English translation task, the bidirectional translators expend significantly more effort on information searches than the English unidirectional translators do. The fact that the bidirectional translators also performed a significantly higher number of searches than the unidirectionals did in the same task may indicate that the two groups have different needs with regard to uncertainty management. The amount of effort that the bidirectional translators invested into their L2 translation task was apparently considered necessary as otherwise they would not have expended it. The aspect of uncertainty management and risk mitigation is discussed in Section 9.2. For the other four indicators for translator effort expended in the German-English task and all five indicators in the English-German task, the statistical analyses did not show any significant differences between the translator groups.

In general, there do not seem to be any differences between the bidirectional and the unidirectional translators with regard to translator effort. It is difficult to compare this finding with those of other studies, such as the ones mentioned in Section 3.5, as they only had bidirectional translators as participants and therefore investigated directionality within that group (i. e. Ferreira 2012 and 2014; Fonseca 2015; Whyatt, Kościuczuk and Turski 2017; da Silva et al. 2017). If I only considered the results for the bidirectional translators in my study, they would support da Silva et al.'s (2017) finding that the translators' effort expended in L2 translation compared to L1 translation shows similarities and differences depending on the indicators investigated. Just as in this study, da Silva et al. (2017, 123) suggested that some results may be related to differences between the task requirements, to which the translators react by expending various amounts of effort.

In addition to the three research questions discussed above, the potential relations between directionality and characteristics of professional translation were investigated by triangulating the results of the five indicators for translation effort with productivity and professional experience for the two participant groups of bidirectional and unidirectional translators under the four conditions bidirectional into L2, bidirectional into L1, English unidirectional into L1 and German unidirectional into L1 (Chapter 8). The triangulation was done through correlations, which suggested some general implications as follows:

- Expending a high amount of effort for character production, revision and information search may have a detrimental effect on productivity.
- Expending a high amount of coordination effort may result in higher productivity, since coordination effort is related to the

translator's switching between the source and the target text in long pauses and therefore to monitoring activities.
- Effort for translation as measured from gaze behaviour did not seem to be related to either productivity or professional experience.
- Extensive professional experience seems to be related to lower effort for character production, revision and information search. In addition, professional experience seems to be positively related to productivity, which is in line with the above-mentioned relation between these effort indicators and productivity. However, for the group of unidirectional English L1 translators, professional experience does not seem to have a positive effect on productivity. This may indicate that there is still room for higher productivity for some translators, but the unidirectional English translators seemed to have reached the maximum and were the most productive group in this study.

On the whole, the indicators of translator effort that are linked to specific process activities have stronger correlations with other characteristics of professional translation such as productivity and professional experience than the indicator from gaze behaviour. In general and perhaps not surprisingly, the more experienced the translators, the more productive they seem to be.

For the bidirectional translators into L2 and the English unidirectional into L1, product quality was included in the correlations. Results for correlations involving product quality suggest that professional experience also may have a positive effect on the ability to expend just the amount of effort needed to produce translations of good quality.

With regard to revision effort, there was only a weak positive correlation with product quality. This could be interpreted as supporting Lorenzo's finding that it is rather the quality of the revision activity and not revision effort per se that impact on product quality (Lorenzo 2002a, 142).

It seems that the higher amount of information search effort that the bidirectional translators had expended in the German-English task compared to that of the English unidirectional translators resulted in a similar positive correlation with product quality. It seems that for both groups, it was generally *worth the effort* as they produced target texts of comparable and accepted quality. This is in accordance with Jääskeläinen's (1999) as well as Lorenzo's (2002b, 103) findings that those translators who invested the highest effort produced the target texts of highest quality. PACTE also found a correlation between information search measures and product quality, which they interpreted as the translators' efficient use of online resources (Kuznik and Olalla-Soler 2018, 23).

In every translation task, the translator strives to find a balance between speed and quality. To concentrate more on either involves an increase in

effort, which may cause fatigue and will eventually decrease productivity (see e. g. Lacruz 2017, 386). The negative correlation between productivity and product quality in the German-English task suggests that high productivity may not necessarily lead to high product quality. This potential trade-off between speed and quality therefore has to be considered by the translator in every task and generally by translation agencies in their role as intermediaries and by employers in their own interest.

Within the group of bidirectional translators, character production effort was the only effort with a significant difference and therefore is probably related to directionality. It showed a strong to very strong negative correlation with the bidirectional translators' productivity and professional experience (see Chapter 8). The link to productivity is rather obvious, since the effort is related to the production of text. As for its correlation with professional experience, which was measured in number of working years, the more experienced translators may simply handle text production more economically.

On the whole, professional experience seems to have a greater relationship to translator effort and productivity for the bidirectionals than for the unidirectional translators, irrespective of translation direction. With regard to the strength of the correlations, there is more similarity between the German unidirectional translators and the bidirectional translators working into their L1 than between the former and the English unidirectional translators although they all worked into their L1.

Preconceptions about translation into a second language and the resultant products as indicated above and in Chapter 1 Introduction have sometimes been negative enough to suggest that this practice is unprofessional. If this were the case, there should be indications of unprofessional practices in the translators' processes as well as in the products and effort. In this study, I investigated whether L2 translators demonstrate professional behaviour in terms of their activities during their TPs. This is a very task-dependent approach to professionalism compared to wider definitions that include professional ethics and service provision.

My premise is that all of the translators who participated in my study have a comparable level of professional experience, competence, and expertise because they are paid practitioners with appropriate training and are confident enough to participate in a study of this nature, so have a well-developed translatorial self-concept. Differences in their behaviour in terms of process activities in the task at hand could then be interpreted as being related to directionality.

Mean length of writing incidents in terms of characters is consistent for the bidirectionals in both tasks and not different from those of the unidirectional translators. Therefore, it seems that their writing patterns are comparable. On the whole, the groups tended to perform the higher percentage of their revisions during the drafting phase when translating

into their L1. In comparison, the bidirectional translators working into their L2 have a rather balanced proportion of revisions with regard to the drafting and the revision phases. This certainly defies the assumption that revision in the L2 is not possible.

Although the bidirectional translators in my study performed a high number of online searches, there does not seem to have been the type of overuse that has been observed with students of translation (e. g. Massey and Ehrensberger-Dow 2014, 86) since the professional translators produced target texts of good quality.

9.2 Discussion of the study results

For the majority of the descriptors for the process, the product and translation effort, no significant differences within the bidirectional group or between the bidirectional group and the unidirectional groups were found. This may indicate that the majority of descriptors are not related to directionality. It seems that they are rather related to individual translation behaviour, to the task at hand or to whether a translator works in one or two translation directions.

The time that the translators take to perform a translation task does not seem to be a reliable indicator of translator effort or of directionality. I suggest that process duration is only the result of several factors and may not be a strong indicator of the level of translation competence or of translator effort. Whyatt (2018, 73) did not find a statistically significant difference in process duration between her participants (language students, translation students and professional translators). She concluded that "[t]his seems to suggest that translation expertise modulates the ways of information processing rather than the overall task duration. Although the professional translators did not use online resources as much as the trainees and the language students, they obviously used their time differently, possibly to reach a high level of accuracy/quality in their target texts" (Whyatt 2018, 74). Da Silva et al. (2017, 127) did not find a difference in task duration either and advised against using it as an effort indicator.

I concluded that the differences between the translator groups in this study do not seem to be related so much to L2 translators compared to L1 translators but to bidirectional translators compared to unidirectional ones. Potential reasons for this finding may be related to the employment situation and related ergonomic aspects of the three participant groups – the German bidirectional translators, the English unidirectional and the German unidirectional translators.

The bidirectional translators in my study are a heterogeneous group: While some of them do almost all of their translation commissions into

their L2, others have a larger percentage of work into their L1. They also differ in other respects: Five of them are freelancers while only one is a staff translator. This may be an important factor that differentiates them from the unidirectional English group and also to some extent from the unidirectional German translators. As we know from a survey in a study on ergonomics at the translators' workplace, the participating freelancers had significantly more freedom in deciding when to do which commission and worked under significantly less time pressure than staff translators did (Ehrensberger-Dow et al. 2016, 13–14). This could mean that for freelancers, high productivity is not of the same importance as it is for staff translators, whose productivity can easily be calculated and checked by their companies' common use of workflow tools (Ehrensberger-Dow et al. 2016, 13). Admittedly, the translation market has undergone dramatic changes in recent years and younger freelancers seem to work more for translation agencies who act as intermediaries between clients and translators and have the capacity to acquire large commissions. Nevertheless, the bidirectional translators in my study like their work, which is an invaluable source of motivation. As Cronin (1998, 93) put it: "lf there is no *play*, the game of translation is likely to lose its appeal".

With regard to their job situations, the English unidirectional translators are the most homogeneous group as they all are staff translators at the same large international language services provider. Productivity is very important to their company, as is working under time pressure because of tight deadlines. There is also internal control since all translations are pre-processed by the project manager in a CAT tool, which then suggests an approximate duration for the completion of each task. These pre-calculated task durations may also have an effect on the translators' information search behaviour, which they need to reduce to an absolute minimum. The translators are used to working under these conditions and do not seem to resent it.

The German unidirectional translators are a slightly more heterogeneous group in terms of employment situation as five are staff translators at the same company as the English translators and one is a freelancer. In this study, they performed similarly to the bidirectional translators in terms of process and effort but not productivity, which was much higher for the unidirectionals. This may also be related to their being used to working at a high pace. With regard to the strength of relations between the characteristics of professional translation, the unidirectional German translators were more similar to the bidirectional translators working into their L1 than to the English unidirectional translators. The reason for this might be related to features associated with the target language of German.

Differences between the bidirectional and unidirectional translators might also be due to differences in risk mitigation and uncertainty management. For example, my study showed that there is a difference

between the bidirectional and the unidirectional translators' information seeking behaviour. One of the possible interpretations is that the bidirectional translators are more cautious than the unidirectional translators and therefore invest more effort in information seeking. They might be more aware of potential problems and tend to check and double check more often. Switching between translation directions might have entrenched a certain behaviour that is now applied to both directions, possibly because it has proved successful during their professional life and ensured an ongoing flow of commissions. By contrast, the unidirectional English translators were all employees in a large language service provider that at the time of data collection still used to have most of the translations proofread internally. The English translators who participated in my study might therefore have become accustomed to refrain from thoroughly revising their texts and only performed information searches when absolutely necessary.

While the use of reference material is commonly seen as a primary problem indicator based on the categorisation by both Krings (1986, 121) and Jääskeläinen (1999, 166), it could also be seen as an integral part of the translation process and, more specifically, as an appropriate way of managing uncertainty. As a group, the bidirectional translators expended more effort in L2 translation than the unidirectional translators in the German-English task with relation to information search queries. This may be the bidirectionals' approach to mitigating risk and to managing uncertainty. The respective results of their translation effort, the target texts of both groups, then are of similar quality when evaluated and ranked by addressees.

As already suggested by Lorenzo (1999), tolerance of uncertainty may be a trait that is more predominant in L2 translation than in L1 translation. The bidirectional translators deal with the uncertainty about whether the communicative act they prepared the target text for will be successful. Although L1 translators have to cope with the same dilemma, they seem to rely more on a *native speaker authority* – a strong concept that seems to be at play in many areas of our western nation states. Perhaps to compensate, the bidirectional translators deal with this uncertainty by doing more information seeking, which probably contributes to longer times on task and lower productivity. It remains unanswered by this study whether the bidirectional translators are more cautious and more rigorous by nature or whether they have acquired this trait during their experience in translating in both directions.

Managing uncertainty and ambiguity as a skill is considered increasingly important in Cognitive Translation Studies and translation training. In the latest EMT Competence Framework (EMT Board 2017) *personal and interpersonal skills* is one of five competence areas whereas in the first framework (EMT expert group & Gambier 2009), *interpersonal skills* featured under *translation service provision competence* and *personal skills*

were not mentioned at all. This is an indication that additional skills are being recognised as required in translation training since some of the others may be covered adequately by neural machine translation (NMT) in the future.

Based on the results of this study, bidirectional translators may have to expend more effort on certain aspects of their tasks to produce the same quality as unidirectional translators at the cost of lower productivity but perhaps fun is a positive factor that also has to be considered in the equation. As is common in an economic context, it probably makes sense to consider what the translator receives in return to expending effort in the task. As an immediate payback, this could be emotions such as satisfaction, a sense of achievement and pride, or relief. Affirmative emotions probably have a positive effect on recovery time. On the one hand, extreme emotional states may compete for cognitive resources, while on the other hand, positive work attitude may help with managing cognitive load and provide further motivation. Effort may therefore come in different shapes comparable to stress. While *eustress* is said to have a positive effect on performance in translation, *distress* has the opposite effect (Bayer-Hohenwarter 2009, 196). For example, a challenging topic and text may have the translator expend more effort in comparison when she has a dull, monotonous task at hand but the overall effect is positive.

It may seem that the fun or playful part of translation has little room in the rigid working conditions and with the heavy use of translation tools in settings such as a global LSP belonging to a venture capitalist's conglomerate. However, as Risku, Milosevic and Pein-Weber (2016, 63–64) indicated, creativity in translation is not limited to the level of word play but can include approaches to problem-solving in a larger context. From my own observations, I know that staff translators also have some leeway in dealing with strict time limits and other constraints, which can be seen as creative problem-solving. On the level of organisations, on an even larger scale, we know from personal accounts of translators' career paths that they exploit the benefits of various job situations by switching between freelance and employed positions depending on their current priorities in life.

If some translators exert greater effort than others in order to produce a target text, this means that this effort was probably necessary in their view. In their daily lives as professionals, translators do not have the possibility to compare their effort to that of others. Even for bidirectional translators, comparing their effort in L2 translation to that in L1 translation may be complex. As long as effort and benefit are in balance in their professional careers, there may be no cause for comparison. Benefits may comprise a combination of challenge, motivation, satisfaction and remuneration. In a wider, sociological context, when a translator is considered part of a network, benefits also include the positive notion of her interdependency with other agents (Risku, Rogl and Pein-Weber 2016), the recognition of her

professional status (Sela-Sheffy 2016) and a sense of agency (Abdallah 2010). Translation can thus be a playful activity, full of excitement and creativity and satisfaction as a payback. Research in educational psychology has shown that learner performance is enhanced by a playful setting (Nørgård, Toft-Nielsen and Whitton 2017). The fun factor in translation should not be underestimated. The same applies to job satisfaction and professional pride.

9.3 Limitations of the study

As mentioned in the Section 4.2 in the Chapter *Methodology*, it was not possible to find unidirectional German-English translators with English as their L2 and bidirectional translators with English as their L1. In order to have a so-called full experimental design, these additional groups would have been necessary. While it seems possible that there are many of the latter, though maybe not in German-speaking countries, the former are probably quite rare. This may be the case for a lot of English L1 translators as they are in high demand anyway. Apart from practical constraints with regard to the sampling, a full experimental design would also have reached a level of complexity that would have been difficult to handle by a single person. Another aspect that would have increased the generalisability of results is to have had all the participants perform a number of tasks, which would have allowed for a repeated measures design (Mellinger and Hanson 2016, 7). However, apart from the researcher's resources, this would have stretched the participants' patience in the usability laboratory.

Another potential limitation of the study concerned the definition of what a pause in a TP is and how it should be operationalised. The discussion of both of these issues is ongoing in the research literature (see e. g. Couto-Vale 2017 or Muñoz and Martín 2018). While my approach of *one threshold fits all* has had a long tradition (see Kumpulainen 2016), the notion of individual pause definitions has gained ground lately (e. g. Dragsted 2005, Muñoz and Cardona 2018). Of course, how a pause is defined also depends on the research interest and on the researcher's definition of the translation process as well as on the available data elicitation method.

My definition of pausing may only represent the pause behaviour of certain participants. It is also defined very narrowly: If a translator tends to follow the text she is reading with the cursor or needs to scroll through the document, this movement prevents the keylogger from recording a long pause or interrupts a pause. My pause definition also had an impact on coordination effort as an indicator of translator effort. Coordination effort involved pauses in the keystroke logging during which the translator looked at least once at the source text and at the target text. A lower pause threshold would have allowed the possibility of catching a higher percentage of pauses

and therefore also of calculating a potentially more representative indication of pauses that involve gaze visits to the source text as well as to the target text. In her study on reading patterns during pauses, Kruger (2016) found that 45 % of all pauses in her participants' TPs into L1 involved looking at both the source text and the target text. However, she used a pause threshold of three seconds and did not include the revision phase in her analysis. While Kruger (2016, 34) reported the mean (i. e. 45 % of all pauses devoted to source text -target text coordination), I reported the median per group. The means in my study range from 46 % to 53 % and the overall mean is 49 %, which is not that much higher, although I included the revision phase in which source text -target text mapping is an important activity.

As in pausing in general, the translator's behaviour during reading also impacts the results for coordination effort: If she tends to follow the lines of the text with the mouse cursor or if she scrolls through the text, the KSL program logs these movements and the whole reading activity is cut into shorter segments. Even if the translator keeps her hand on the mouse, she may unintentionally move it while reading and/or thinking. This movement is also logged and therefore may subdivide pauses into units that no longer fall into my pause category. My definition of pausing also had a direct impact on the other three process activities that I investigated. For example, when a translator paused her writing activities for five seconds or less and then continued, this was categorised and counted as one writing incident. However, if she had paused for longer than five seconds while writing, this would have resulted in two writing and one pausing incidents. Therefore, it seems that the definition of a pause was a basic decision that influenced many of the results in this study. However, the same definition was applied to both tasks and all of the groups.

The definition of a pause also had an effect on the amount and length of what I named 'other activities', as reported in Section 5.2.2. The proportion of those other activities was quite high compared to process duration. The effort expanded on those activities are accounted for in the effort indicator of mean fixation duration. In a further analysis or a follow-up study, it would probably be feasible to have additional indicators to account for that effort. One approach might be to determine the overall keyboard activity in relation to the number of characters in the final TT. This was not possible in this study as two different versions of the keystroke logger were used, which did not guarantee comparability of the proposed measure. Most importantly, though, is having several indicators of effort as one indicator cannot provide the same insights as multiple indicators can. When designing a follow-up study, it would also be important to consider that the amount of keyboard activity more likely represents individual than general translation behaviour.

It seems obvious that every method of data elicitation that has been added to the CTS researchers' toolbox has provided additional insights into

translation. The opportunities that have opened up with each new technique have been adopted with enthusiasm and exploited to the maximum. Only in hindsight has the awareness grown that every new method is accompanied by new limitations and that they need to be considered carefully.

The requirements of objectivity were met in this study by applying a combination of well-established, tested and critically discussed data elicitation tools within cognitive translation studies. The overall mixed-method approach facilitated validity as some blind spots in one method could be illuminated by another method. An example are the activities during KSL pauses that could be traced by eye gaze representations in the screen recordings. The issue of reliability of results was addressed by controlling extraneous variables such as the level of professional experience by balancing them in the participant groups.

Transparency was achieved by an in-depth analysis of the data and by the reporting of significant as well as non-significant results. The detailed description of sampling, operationalisations and procedures facilitates repeatability of the study. However, replicability of all aspects of the study are acknowledged to remain an unresolvable issue, a theoretical goal in TS as in any other people-centred research strand. Validity was addressed by measuring what I wanted to measure, under the premise that process activities are manifestations of underlying cognitive activities and effort indicators related to the whole translation process as defined in my study. Product quality was assessed holistically and as a relative and not an absolute measure in order to make the task manageable for the raters, among other goals. Both quantitative and qualitative data was analysed quantitatively whenever possible to allow for comparisons, statistical analyses and correlations. Ecological validity was achieved by designing the experimental situation to be as realistic as possible.

The general best practice recommendation is that only a combination of methods and an abundance of studies will allow us to widen the scope of the field, always with the recognition of never being able to provide a complete picture of the cognitive aspects of translation.

9.4 Relevance of the study and proposition for reconceptualisation

This study's uniqueness and advantages lie in its multi-method approach, its inclusion of a comparison group of L1 translators, its triangulation of process and product data and its investigation of translator effort that is correlated with other characteristics of professional translation such as productivity and professional experience.

The present study has substantiated the complexity of directionality and of its effects on the translators' processes, products and effort for translation.

As I had bidirectional translators working into their L2 and their L1 as well as unidirectional translators working into one of the language versions, it was possible to consider the same indicator from different perspectives. A case in point is process duration. If I had only compared the bidirectionals translating into their L2 to the English unidirectional translators working into their L1, process duration would have been identified as an indicator of directionality. However, the additional comparisons revealed that the bidirectional translators were also significantly slower when working into their L1 than the German unidirectional translators were. Moreover, there was no significant difference between the bidirectionals when working into their L2 and into their L1. As a consequence, process duration does not seem to be a suitable indicator for directionality. The same can be said for its use as an indicator of translator effort.

Without delving into the ongoing discussion of competence versus levels of expertise (see e.g. Shreve 2006 or Muñoz 2014), experience in translation may need to be operationalised in some way other than simply stating the years of professional practice. This is also a desideratum of other studies involving professional translators, such as that of Förster Hegrenæs (2018, 249), as years of experience does not seem to appropriately reflect the level of task entrenchment. As the comparison of the effort scores per translation direction for the bidirectional translators showed, their self-reported workload in each direction may have had an impact on translator effort. A high level of entrenched or routine behaviour may lead to a different approach to managing uncertainty in the translation process as manifested in revision and information search activities. Even exposure to and immersion in source and target languages could be considered to impact professional practice, as has been suggested by Chmiel (2016, 288).

Another issue concerns the status of the activity of information searches in cognitive translatology. If information seeking is seen as an integral part of the translation task and documentary resources as part of the environment needed for human cognition to happen, then it may be reasonable to reconceptualise this activity in cognitive translatology. The use of information resources has probably far too often been associated with controlled processes or intentional actions. The approach to conceptualise information seeking as a situated action (Martín 2008, 20) could free it from being considered as a sequence of intentional actions that then can be qualified as strategic or non-strategic behaviour (Göpferich 2012), as effective and efficient or the contrary.

Among other scholars, Whyatt, Kajzer-Wietrzny and Stachowiak (2017, 151) suggest that information searching may distract the translators from the actual task (i.e. from producing a TT). This means that while acknowledging that information searching is "a part of translators' real-life performance" (Whyatt, Kajzer-Wietrzny and Stachowiak 2017, 151), it is

considered to be a distraction. However, the translator may feel that without going to an online resource, she would not be able to continue with her task. Hence, information searches should be considered an integral part of the translation process and therefore also of the cognitive processes involved.

It seems that since incidents of information searches have been treated in the past as interruptions, they have been investigated as activities separate from the actual translation process (see e. g. B. Nord 2002, 115 or Gough 2015, 62). In that approach, efficacy of information searches then has, for example, been defined as quickly finding an expression that was displayed in one of the resources and directly including it in the TT. In addition, the solution needed to be correct in the eyes of the evaluator. The production of inacceptable solutions or the inefficient use of time to find a solution have often been related to the lack of a strategic approach to problem solving (Göpferich 2012, 241; Hunziker Heeb 2012, 184; B. Nord 2002, 115; PACTE 2011, 339; Prassl 2011, 44). While creating a macrostrategy that guides the task is not questioned at this point, it may be reasonable to relate information searches to cognitive load or, if performed without a satisfactory result, even to cognitive overload. It seems that deciding on the most effective way to do a search is far easier and very different for an observer than for the translator actually performing the search within a task. Actually, the two cognitive processes involved cannot be compared and therefore, the observer should be cautious when making inferences about the translator's decisions. Non-efficient use could be seen as an indicator of cognitive overload (e. g. Ehrensberger-Dow and Hunziker Heeb 2016, 81) rather than of the lack of a translation strategy.

The use of information resources certainly has an effect on task duration and probably also on cognitive resources, as switching between tasks is considered taxing (Whyatt 2018, 66). It seems questionable in what sense information searches distract from the actual translation task. If only the source text reading and target text writing parts are considered to be translation, then information searching is interrupting that activity. However, if all the activities involved in the task are conceptualised as translation, then information seeking has to be treated as an integral part of it.

Differences emerged not so much between L1 and L2 translation but between unidirectional staff and bidirectional freelance translators. Therefore, directionality may still be too narrow a concept. Professional practice has to be considered. For a bidirectional translator, L2 translation cannot be separated from L1 translation. Just as the languages a person speaks influence each other, so do the translation directions a translator works in influence each other. There may therefore be some potential in conceptualising bidirectionality not as an absolute deviation from unidirectionality but as a variant of multi-directionality. The actual practice could give an indication of dominant translation direction, maybe comparable

to current conceptualisations of multilingualism and efforts in determining individuals' dominant language(s) for specific purposes and in specific domains. As research in neuropsychology suggests that the brain's plasticity and entrenched patterns of behaviour, to which I would count translation direction, help handling complex tasks (Elmer, Hänggi and Jäncke 2014), multi-directionality may not necessarily be an impossible practice in terms of available cognitive resources. Admittedly, widening the discussion from uni- to multi-directionality does not simplify matters but it may help call into question notions of directionality that have led to unjustified preconceptions with discriminatory effects for some translation practitioners.

I have deployed a user-centred approach to product quality evaluation as I wanted readers to assess the TTs similar to those that would naturally do so in real life and under the conditions that they might do so. Apart from the translators themselves and the clients who pay for the translation, the text users have the greatest interest in good product quality in a professional setting. I think this justifies my approach, which could be used in comparable contexts. I suggest investigating whether such an approach to evaluating professional translators' products solves some of the issues of process quality not matching product quality that Jääskeläinen (2016) discusses.

9.5 Suggestions for further research

Apart from the suggestions that could be made to improve the study discussed in the previous sections, such as larger number of participants, a workplace study, other language combinations, longer source texts, a different approach to assess product quality, it would be worth venturing into other fields. Using this study as a starting block, further research could cover motivational aspects of L2 translation training, evaluators' attitudes towards L2 translation and the reuse of the translator effort indicators.

To investigate what translation students actually need to foster their L2 translation skills and to strengthen their self-confidence, classroom studies with active participation of students and lecturers are necessary. These could then be combined with follow-up studies during professional life. Professional L2 translators could comment on needs that had been catered for during training and needs that in retrospect would have been helpful. Teachers with positive attitudes towards L2 translation, not only into English but in general, are probably a decisive factor in successful training. Enhancing teachers' knowledge of what translation competence is comprised of and how it can be fostered in their very own classroom could contribute to empowering their students to pursue career paths that include translation into and out of their L1.

A follow-up study could investigate translation evaluators' attitudes toward L2 translation and correlate them with the results of the product evaluation. Professional revisers, translation teachers and text users could be recruited as evaluators. For the study, all target texts would need to be randomly assigned to two groups: either the one presented to the evaluators as produced by translators into their L2 or the one presented as produced by translators into their L1. Their evaluations would then be compared to the evaluations by raters who were not informed about the translation direction. Such a study would provide insights into relations between the attitude towards L2 translation and product evaluation. It would also inform on the different group of evaluators' potential to identify the translation direction.

The indicators of translator effort that were proposed in this study could be used to investigate other translation aspects than directionality. In fact, they will be used in an ongoing research project on cognitive load in interpreting and translation (CLINT 2018). The study investigates whether the translation of texts written in non-standard English involves more translator effort than of texts in standard English. In parallel, it will be investigated whether the phenomenon has an impact on interpreters' performance and other multilinguals' text comprehension. Results from Cognitive Translation Studies and Interpreting Studies will be triangulated with results from neuropsychology, which uses neurophysiological markers such as brain responses to measure cognitive load.

9.6 Potential implications of the study's findings

The transferability of this study's findings is promoted by presenting potential implications for translation training and translation service providers. In addition, suggestions for further-reaching changes in attitudes and preconceptions are made.

Any further empowerment of translation students could draw on the finding that there are many similarities between translation directions without neglecting the few differences. In my opinion, teaching institutions have a certain duty to prepare their students for the reality of the market, to enable them to find work and to assist them towards the goal of becoming reflective practitioners. As there is work available in translating into English in particular, many graduates could find work translating into their L2 or L3. More specifically, the results of this study suggest providing students with approaches adjusted to the translation direction. This could be done, for example, by investing in tailored information search practice of electronic resources, choosing adequate resources or even compiling them themselves (for an example on using ad-hoc corpora see Rodríguez-Inés 2014). Students also need to be introduced to effective self-revision in L2

and to the possibilities of cooperating with revisers, who may or may not be L1 speakers of the target language.

A powerful technology that needs to be considered in translation training as it seems to be here to stay, is NMT. Students need to find ways to optimise their use of NMT while at the same time omitting its pitfalls and being *ahead of the game*. Some NMT tools that are commonly accessible online seem to provide translations of an astonishingly acceptable quality in certain aspects, especially if English is one of the working languages. One aspect is the syntactically and grammatically smooth surface structures of the target texts under the provision that the source texts are also of this quality. The controlled use of NMT may free translators from worrying about general text mechanics such as grammar, spelling or punctuation, which may be beneficial to L2 translators in particular. Instead, the translators could direct their cognitive resources to more complex problems, such as errors of content, sense or logic. To spot and correct such errors really calls for advanced translation skills. In that sense, efficiently using NMT for translation purposes may need training institutions and individual translators to make modifications in text production skills towards more revision skills (cf. Mossop 2014). The translator-reviser needs to focus on text coherence and cohesion and on text logic in general. In addition, translators may need to focus on pre-editing source texts. As this requires text production skills in the source language, bidirectional translators are probably in the lead. Up to now, NMT has worked on the sentence level only, something that may never be overcome, as it apparently would require substantially more computing power to even take the context of the adjacent sentences let alone the entire document into account (Läubli, Sennrich and Volk 2018). Translators who consider working with NMT less appealing may instead find work as what is sometimes referred to as transcreators (cf. Pedersen 2017), where creativity and meeting clients' specific requirements and introducing recipient orientation are in high demand. Of course, the appropriation of NMT for translators also opens a whole new area for professional development courses and for research.

As my study has shown that product quality is not an issue for professional translators accustomed to translating in both directions, teaching should focus on the translation process and on getting experience by performing authentic tasks. This could help replace a text-genre approach that conceptualised translation as mainly a comparative linguistics task based on input from genre studies and contrastive rhetoric (cf. Beeby 2003) and break the circle of reproducing outdated curriculum ideologies (Kearns 2012, 25). Rodríguez-Inés and Fox (2018), for example, reported on the successful introduction of a competence-based, learner-centred, process-oriented curriculum for L2 translation teaching. Following Schäffner's (2000) argumentation that translation competence and L2 language skills can develop in parallel,

translation into L2 could generally be taught at undergraduate level already. Overall, curriculum designers could benefit from the experience of training institutions where preparing the students for the market's need of L2 translation into English has been business as usual for a long time.

In addition, performance evaluations on the Master's level could increasingly involve potential target text users. This could foster the students' self-confidence as they would learn to focus on the users' needs and thereby build a macrostrategy that would free them from sticking too closely to the source text. On a different level, it would also prepare students to work as technical translators as often it makes sense to collect and consider the needs of user groups as early as possible in the translation process as an opportunity to manage potential risks (cf. Suojanen et al. 2015).

Generally, L2 translation teaching could vastly benefit from information on the actual practice as provided by professional L2 translators, or expressed more accurately, by professional bidirectional translators. One possibility would be to invite professional bidirectional translators to talk about their work and experience and thereby provide role models and examples of career paths to translation students. Another possibility, or rather necessity in my opinion, would be to employ teachers with professional experience in bidirectional translation. I think their positive mind-set and practical approach could sustainably empower L2 translation students. Therefore, it is high time for the call for L2 speakers of the target language to be translation teachers is heard and acted upon.

What has been said with regard to the training of translation students also applies to professional translators. There is much scope for professional development courses directed at the specific needs of L2 translation. As already mentioned above on the implications for translation training, the use of NMT in professional translation by translators as well as pre- and post-editors opens new career opportunities for L2 translators into English. The recently developed International Standard on post-editing contains specifications on pre- and post-editing and competences almost identical to those needed by translators, without prescribing the status of the performer's source and target languages (ISO 18587, 2017).

Conceptualising translators as potential bidirectional or even multi-directional translators may have a positive effect on knowledge transfer between translation directions. In my opinion, this approach could also be effective for translation students with A-C-C language versions as a preparation for their lives as professionals.

In the light of the finding that directionality has no significant effect on the quality of the translation products in this study, bidirectional translators should make use of their bidirectionality and promote what clients probably already assume: Professional translators are able to work into and out of their L1 and L2 with no difference in quality.

Another aspect could be collaborations between L1 speakers of the target and of the source languages. Whatever the status of the target language speaker, translators could benefit from collaborating and not treating L1 speakers of the target or source language as ultimate authorities on language matters, since translation is about producing a text appropriately suited for a new audience. Envisioning the audience as a mixture of L1 to Ln speakers of English whose common denominator is that they are users of the target text, may open new perspectives: Translators may want to focus on aspects such as readability and comprehensibility rather than aiming for linguistic perfection as an unsuitable and unattainable goal, irrespective of translation direction.

Professional translators who themselves would never translate into their L2 should reassess their attitudes and preconceptions towards L2 translation and grant others the cognitive capacity, the skills and fun to practice L2 translation in a professional way. Rather than excluding L2 translators from becoming members of professional associations and thereby ignoring translational practices in the market, other admission criteria should be considered in order to strengthen the profession's status and position in the market. The necessity for large and small players in the translation market to introduce more sensible quality criteria than 'native speakerhood' in the target language is evident. Such criteria could be based on the regulations for quality assurance specified by the International Standard ISO 17100 (2015) or on a more user-centred approach. The most straightforward criteria may be to ensure that commissioned translators have formal translation training or accreditation.

For the time being, no other language seems to be able to substitute for English as a lingua franca. Thus, the demand for translation into English is a given and will continue to be so in the foreseeable future. With regard to the status of the languages we speak, I have the impression that growing up with a single L1 is a situation that is on the verge of extinction. Many people already are bilingual or trilingual without being equally fluent in all their languages in all life situations. Thus, the labelling of one's languages and the attempt of a strict separation into L1, L2, and Ln seems to make less and less sense. With regard to translation, I would therefore opt for a pragmatic approach that allows translation students and professionals to acquire the competences that enable them to obtain work in whatever language combinations and translation directions they feel comfortable with and to adapt to market needs.

10 Bibliography

Abdallah, K. (2010). Translators' agency in production networks. In T. Kinnunen & K. Koskinen (Eds.), *Translators' agency* (pp. 11–46). Tampere: Tampere University Press.

Alves, F. (Ed.). (2003). *Triangulating translation: Perspectives in process oriented research* (Vol. 45). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.45

Alves, F., & Gonçalves, J. L. (2013). Investigating the conceptual-procedural distinction in the translation process. *Target*, *25*(1), 107–124. https://doi.org/10.1075/target.25.1.09alv

Alves, F., Koglin, A., Mesa-Lao, B., García Martínez, M., de Lima Fonseca, N. B., de Melo Sá, A., Gonçalves, J. L., Sarto Szpak, K., Sekino, K., Aquino, M. (2016). Analysing the impact of interactive machine translation on post-editing effort. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New directions in empirical translation process research* (pp. 77–93). Cham: Springer. https://doi.org/10.1007/978-3-319-20358-4

Alves, F., & Vale, D. (2009). Probing the unit of translation in time: Aspects of the design and development of a web application for storing, annotating, and querying translation process data. *Across Languages and Cultures*, *10*(2), 251–273. https://akjournals.com/view/journals/084/10/2/article-p251.xml

André, J. S. (2010). Lessons from Chinese history: Translation as a collaborative and multi-stage process. *TTR : Traduction, Terminologie, Rédaction*, *23*(1), 71–94. https://doi.org/10.7202/044929ar

Aneja, G. A. (2016). Rethinking nativeness: Toward a dynamic paradigm of (non)native speakering. *Critical Inquiry in Language Studies*, *13*(4), 351–379. https://doi.org/10.1080/15427587.2016.1185373

Apfelthaler, M. (2013). Directionality research in translation and interpreting studies (1/2): A shorter than short history. Retrieved 1 February 2014, from http://cogtrans.net/blog.htm

Apfelthaler, M. (2014). Stepping into others' shoes: a cognitive perspective on target audience orientation in written translation. *MonTI. Monografías de Traducción e Interpretación*, 303–330. https://doi.org/10.6035/MonTI.2014.ne1.10

Asadi, P., & Séguinot, C. (2005). Shortcuts, strategies and general patterns in a process study of nine professionals. *Meta*, *50*(2), 522–547. https://doi.org/10.7202/010998ar

Audeoud, M., & Haug, T. (2013). 'Directionality Effect' beim Gebärdensprachdolmetschen – Welche Dolmetschrichtung präferieren Gebärdensprachdolmetscher in der Schweiz? Teil 1. *Das Zeichen*, *27*(93), 124–136.

Balling, L. W., & Hvelplund, K. T. (2015). Design and statistics in quantitative translation (process) research. *Translation Spaces*, *4*(1), 169–186. https://doi.org/10.1075/ts.4.1.08bal

Balling, L. W., Hvelplund, K. T., & Sjørup, A. (2014). Evidence of parallel processing during translation. *Meta*, *59*(2), 234–259. https://doi.org/10.7202/1027474ar

Bayer-Hohenwarter, G. (2009). Methodological reflections on the experimental design of time-pressure studies. *Across Languages and Cultures*, *10*(2), 193–206. https://akjournals.com/view/journals/084/10/2/article-p193.xml

Bayer-Hohenwarter, G. (2011). Kreativität in populärwissenschaftlichen und instruktiven Texten im Vergleich: Kein Raum für Kreativität beim Übersetzen von instruktiven Texten? *Trans-Kom*, *4*(1), 49–75.

BDÜ. (2017). Bundesverband der Dolmetscher und Übersetzer e. V. Membership database. Retrieved 21 September 2017, from https://search.bdue.de/

Beeby, A. (2003). Genre literacy and contrastive rhetoric in teaching inverse translation. In D. Kelly, A. Martin, M.-L. Nobs, D. Sánchez, & C. Way (Eds.), *La direccionalidad en traducción e interpretación: perspectivas teóricas, profesionales y didácticas* (pp. 155–166). Granada: Atrio.

Bittner, H. (2014). Das Problem der Subjektivität bei der Beurteilung von Übersetzungen. In I. Bartoszewicz & A. Małgorzewicz (Eds.), *Translationsforschung: Methoden, Ergebnisse, Perspektiven* (Vol. 5, pp. 21–35). Wrocław: Oficyna Wydawnicza ATUT.

Buchweitz, A., & Alves, F. (2006). Cognitive adaptation in translation: an interface between language direction, time, and recursiveness in target text production. *Letras de Hoje*, *41*(2), 241–272.

Campbell, S. (1998). *Translation into the second language*. New York: Addison Wesley Longman.

Campbell, S. (2000). Choice network analysis in translation research. In M. Olohan (Ed.), *Intercultural faultlines* (pp. 29–42). Manchester: St. Jerome.

Carey, D. (2016). English will remain the working language of the EU after Brexit. *The Irish Times*. Retrieved from https://www.irishtimes.com/opinion/english-will-remain-the-working-language-of-the-eu-after-brexit-1.2725840

CESE. (2017). Cognitive load theory: Research that teachers really need to understand. Sydney: Centre for Education Statistics and Evaluation. Retrieved from https://www.cese.nsw.gov.au//images/stories/PDF/cognitive-load-theory-VR_AA3.pdf

Chang, C., & Schallert, D. L. (2007). The impact of directionality on Chinese/English simultaneous interpreting. *Interpreting*, *9*(2), 137–176. https://doi.org/10.1075/intp.9.2.02cha

Chang, V. C.-Y. (2011). Translation directionality and the revised hierarchical model: An eye-tracking study. In S. O'Brien (Ed.), *Cognitive explorations of translation* (pp. 154–174). London: Continuum.

Chmiel, A. (2016). Directionality and context effects in word translation tasks performed by conference interpreters. *Poznan Studies in Contemporary Linguistics*, *52*(2). https://doi.org/10.1515/psicl-2016-0010

Christoffels, I., De Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, *54*, 324–345. https://doi.org/10.1016/j.jml.2005.12.004

Colina, S. (2008). Translation quality evaluation: Empirical evidence for a functionalist approach. *The Translator*, *14*(1), 97–134. https://doi.org/10.1080/13556509.2008.10799251

Colina, S. (2009). Further evidence for a functionalist approach to translation quality evaluation. *Target*, *21*(2), 235–264. https://doi.org/10.1075/target.21.2.02col

Conde, T. (2011). Translation evaluation on the surface of texts: a preliminary analysis. *Jostrans – The Journal of Specialised Translation*, *15*, 69–86.

Council of Europe. (2018). Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). Retrieved 13 December 2018, from https://www.coe.int/en/web/common-european-framework-reference-languages

Couto-Vale, D. (2017). What does a translator do when not writing? In S. Hansen-Schirra, O. Czulo, & S. Hofmann (Eds.), *Empirical modelling of translation and interpreting* (pp. 209–237). Berlin: Language Science Press. https://doi.org/10.5281/zenodo.1090970

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks: Sage.

Cronin, M. (1998). Game theory and translation. In M. Baker (Ed.), *Routledge encyclopedia of translation studies* (pp. 91–92). London: Routledge.

CTP. (n. d.). Capturing translation processes (CTP). Retrieved 2 April 2014, from https://www.zhaw.ch/en/linguistics/research/translation-studies/capturing-translation-processes-ctp/

da Silva, I. A. L., Alves, F., Schmaltz, M., Pagano, A., Wong, D., Chao, L., Leal, A. L. V., Quaresma, P., Garcia, C., da Silva, G. E. (2017). Translation, post-editing and directionality. A study of effort in the Chinese-Portuguese language pair. In A. L. Jakobsen & B. Mesa-Lao (Eds.), *Translation in Transition: Between cognition, computing and technology* (Vol. 133, pp. 107–134). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.133

da Silva, I. A. L., & Pagano, A. S. (2017). Cognitive effort and explicitation in translation tasks. In S. Hansen-Schirra, O. Czulo, & S. Hofmann (Eds.), *Empirical modelling of translation and interpreting*. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.1090962

Daems, J., Carl, M., Vandepitte, S., Hartsuiker, R., & Macken, L. (2016). The effectiveness of consulting external resources during translation and post-editing of general text types. In *New directions in empirical translation process research* (pp. 111–133). Cham: Springer. https://doi.org/10.1007/978-3-319-20358-4_6

De Groot, A. M. B. (1997). The cognitive study of translation and interpretation: Three approaches. In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 25–56). Thousand Oaks: Sage.

De Groot, A. M. B., & Christoffels, I. (2006). Language control in bilinguals: Monolingual tasks and simultaneous interpreting. *Bilingualism: Language and Cognition*, *9*(2), 189–201. https://doi.org/10.1017/S1366728906002537

den Dekker, A. (2017). *A survey-based study on the opinions and attitudes concerning professional translation into a non-mother tongue in the Swiss translation market*. Master's thesis. Zurich University of Applied Sciences.

Derwing, T., Rossiter, M., & Ehrensberger-Dow, M. (2002). 'They speaked and wrote real good': Judgements of non-native and native grammar. *Language Awareness*, *11*, 84–99. https://doi.org/10.1080/09658410208667048

Directorate General for Translation. (2009). Wanted: English language translators. MEMO/09/76. DGT. Retrieved from http://ec.europa.eu/dgs/translation/publications/index_en.htm

Doherty, S. (2017). Issues in human and automatic translation quality assessment. In D. Kenny (Ed.), *Human issues in translation technology* (pp. 131–148). London ; New York: Routledge, Taylor & Francis Group.

Dragsted, B. (2005). Segmentation in translation. Differences across levels of expertise and difficulty. *Target*, *17*(1), 49–70. https://doi.org/10.1075/target.17.1.04dra

Dragsted, B. (2012). Indicators of difficulty in translation: Correlating product and process data. *Across Languages and Cultures*, *13*(1), 81–98. https://doi.org/10.1556/Acr.13.2012.1.5

Ehrensberger, S. (2014). TransVis – a translation process visualizer tool. Retrieved 7 July 2016, from https://github.com/sykaeh/TransVis

Ehrensberger-Dow, M. (2014). Challenges of translation process research at the workplace. *MonTI Monographs in Translation and Interpreting*, *Special Issue 1* (Journal Article), 355–383. https://doi.org/10.6035/MonTI.2014.ne1.12

Ehrensberger-Dow, M. (2017). An ergonomic perspective of translation. In J. W. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 332–349). Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119241485.ch18

Ehrensberger-Dow, M. (2018). Process research. In L. D'hulst & Y. Gambier (Eds.), *A History of modern translation knowledge: Sources, concepts, effects* (pp. 293–300). Amsterdam: John Benjamins.

Ehrensberger-Dow, M., & Hunziker Heeb, A. (2016). Investigating the ergonomics of a technologized translation workplace. In R. Muñoz Martín (Ed.), *Reembedding translation process research* (Vol. 128, pp. 69–88). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.128.04ehr

Ehrensberger-Dow, M., Hunziker Heeb, A., Jud, P., & Angelone, E. (2017). Insights from translation process research in the workplace. In D. Perrin & U. Kleinberger (Eds.), *Doing applied linguistics: enabling transdisciplinary communication* (pp. 116–123). Berlin: De Gruyter. https://doi.org/10.1515/9783110496604-014

Ehrensberger-Dow, M., Hunziker Heeb, A., Massey, G., Meidert, U., Neumann, S., & Becker, H. (2016). An international survey of the ergonomics of professional translation. *ILCEA Revue de l'Institut Des Langues et Cultures d'Europe, Amérique, Afrique, Asie et Australie*, *27*. https://doi.org/10.4000/ilcea.4004

Ehrensberger-Dow, M., & Massey, G. (2008). Exploring translation competence by triangulating empirical data. In R. Harvey, I. Hempen, & M. Sieradzki (Eds.), *Experiments in translation* (Vol. 16, pp. 1–20). Norwich: University of East Anglia.

Ehrensberger-Dow, M., & Massey, G. (2013). Indicators of translation competence: Translators' self-concepts and the translation of titles. *Journal of Writing Research*, *5*(1), 103–131. https://doi.org/10.17239/jowr-2013.05.01.5

Ehrensberger-Dow, M., & O'Brien, S. (2015). Ergonomics of the translation workplace: Potential for cognitive friction. *Translation Spaces*, *4*(1), 98–118. https://doi.org/10.1075/ts.4.1.05ehr

Elmer, S., Hänggi, J., & Jäncke, L. (2014). Processing demands upon cognitive, linguistic, and articulatory functions promote grey matter plasticity in the adult multilingual brain: Insights from simultaneous interpreters. *Cortex*, *54*, 179–189.

EMT Board. (2017). *European Master's in translation: Competence framework 2017*. Retrieved from https://ec.europa.eu/info/sites/info/files/emt_competence_fwk_2017_en_web.pdf

EMT expert group. (2009). *Competences for professional translators, experts in multilingual and multimedia communication*. Brussels.

EN. (2006). British Standard EN 15038: Translation services — Service requirements.

Englund Dimitrova, B., & Tiselius, E. (2014). Retrospection in interpreting and translation: Explaining the process? *MonTI Monographs in Translation and Interpreting*, *Special Issue 1*, 177–200. http://dx.doi.org/10.6035/MonTI.2014.ne1.5

Enríquez Raído, V. (2014). *Translation and web searching*. New York: Routledge.

European Commission. (2010). *Translation at the European Commission – a history*. Luxembourg: Office for Official Publications of the European Communities. https://doi.org/10.2782/16417

European Commission. (2014). *Translation and multilingualism*. Luxembourg: Publications Office of the European Union. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/e0770e72-afa1-4971-8824-6190512537dc/language-en

Ferreira, A. (2010). *Direcionalidade em tradução: uma investigação do processamento cognitivo de tradutores profissionais em tradução direta e inversa no par lingüístico inglês-português*. Master's thesis. Universidade Federal de Minas Gerais.

Ferreira, A. (2012). Investigando o Processamento Cognitivo de Tradutores Profissionais em Tradução Direta e Inversa no Par Linguístico Inglês-Português. *Cadernos de Tradução*, *1*, 73–92. https://doi.org/10.5007/2175-7968.2012v1n29p73

Ferreira, A. (2014). Analyzing recursiveness patterns and retrospective protocols of professional translators in L1 and L2 translation tasks. *Translation and Interpreting Studies*, *9*(1), 109–127. https://doi.org/10.1075/tis.9.1.06fer

Ferreira, A., Schwieter, J. W., Gottardo, A., & Jones, J. (2016). Cognitive effort in direct and inverse translation performance: Insight from eye-tracking technology. *Cadernos de Tradução*, *36*(3), 60–81. https://doi.org/10.5007/2175-7968.2016v36n3p60

Flaounas, I., Ali, O., Lansdall-Welfare, T., De Bie, T., Mosdell, N., & Cristianini, N. (2013). Research methods in the age of digital journalism. *Digital Journalism*, *1*(1), 102–116. https://doi.org/10.1080/21670811.2012.714928

Flesch, R. (n. d.). How to write plain English. Retrieved 10 October 2016, from http://www.mang.canterbury.ac.nz/writing_guide/writing/flesch.shtml

Fonseca, N. B. de L. (2015). Directionality in translation: Investigating prototypical patterns in editing procedures. *Translation & Interpreting*, *7*(1), 111–125. http://trans-int.org/index.php/transint/article/viewFile/372/187

Förster Hegrenæs, C. (2018). *Translation competence development and the distribution of cognitive effort: an exploratory study of student translation behavior*. Doctoral dissertation. NHH-Norwegian School of Economics.

Gile, D. (2005). Directionality in conference interpreting: a cognitive view. In R. Godijns & M. Hinderdael (Eds.), *Directionality in interpreting: The 'retour' or the native?* (pp. 9–26). Ghent: Communication & Cognition.

Gile, D. (2009). *Basic concepts and models for interpreter and translator training: Revised edition* (2nd ed., Vol. 8). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.8

Gile, D. (2016). Experimental research. In C. V. Angelelli & B. J. Baer (Eds.), *Researching Translation and Interpreting* (Vol. 1, pp. 220–228). New York; Oxon: Routledge.

Göpferich, S. (2008). *Translationsprozessforschung. Stand – Methoden – Perspektiven*. Tübingen: Narr.

Göpferich, S. (2009). Towards a model of translation competence and its acquisition: the longitudinal study TransComp. In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), *Behind the Mind. Methods, Models and Results in Translation Process Research* (Vol. 37, pp. 17–43). Copenhagen: Samfundslitteratur.

Göpferich, S. (2012). Tracing strategic behaviour in translation processes: Translation novices, 4th-semester students and professional translators compared. In S. Hubscher-Davidson & M. Borodo (Eds.), *Global trends in translator and interpreter training* (pp. 240–266). London: Continuum.

Göpferich, S. (2013). Translation competence. Explaining development and stagnation from a dynamic systems perspective. *Target*, *25*(1), 61–76.

Göpferich, S. (2017). Cognitive functions of translation in L2 writing. In J. W. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 402–422). Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9781119241485.ch22

Göpferich, S., Bayer-Hohenwarter, G., Prassl, F., & Stadlober, J. (2011). Exploring translation competence acquisition: Criteria of analysis put to the test. In S. O'Brien (Ed.), *Cognitive explorations of translation* (pp. 57–85). London: Continuum.

Gough, J. (2015). *The patterns of interaction between professional translators and online resources*. Doctoral dissertation. University of Surrey. Retrieved from http://epubs.surrey.ac.uk/id/eprint/813254

Grabowski, J., Becker-Mrotzek, M., Knopp, M., Jost, J., & Weinzierl, C. (2014). Comparing and combining different approaches to the assessment of text quality. In D. Knorr, C. Heine, & J. Engberg (Eds.), *Methods in Writing Process Research* (Vol. 13, pp. 147–165). Frankfurt am Main: Peter Lang.

Grauwinkel, G., & Neunzig, W. (2008). Fundamentos de un modelo para la traducción especializada inversa y propuesta de validación. In L. Pegenaute, J. DeCesaris, M. Tricás, & E. Bernal (Eds.), *Actas del III Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación.La traducción del futuro: mediación lingüístca y cultural en el siglo XXI. Barcelona 22–24 de marzo de 2007*. (Vol. 2, pp. 213–224). Barcelona: PPU.

Groethuysen, C. (2015). Teaching L2 translation for professional purposes. Presented at the Foreign-language competence for future language professionals: Reassessing market needs and training programmes, Università degli Studi di Trieste.

Grosman, M., Kadrić, M., Kovačič, I., & Snell-Hornby, M. (Eds.). (2000). *Translation into non-mother tongues: in professional practice and training* (Vol. 2nd). Tübingen: Stauffenburg.

Gross, R. (2000). Übersetzen in die fremde Sprache? Leicht gesagt! - Wie man's aber vielleicht trotzdem noch unterrichten kann. In M. Grosman, M. Kadrić, I. Kovačič, & M. Snell-Hornby (Eds.), *Translation into non-mother tongues: in professional practice and training* (pp. 175–183). Tübingen: Stauffenburg.

Hagemann, S. (2016). (Non-)professional, authentic projects? Why terminology matters. In Don Kiraly (Ed.), *Towards authentic experiential learning in translator education* (pp. 33–51). Göttingen: V&R unipress.

Halverson, S. (2010). Cognitive translation studies. Developments in theory and method. In G. M. Shreve & E. Angelone (Eds.), *Translation and Cognition* (pp. 349–369). Amsterdam: John Benjamins. https://doi.org/10.1075/ata.xv.18hal

Halverson, S. (2017). Multimethod approaches. In J. W. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 193–212). John Wiley & Sons. https://doi.org/10.1002/9781119241485.ch11

Hansen, G. (2003). Controlling the process. Theoretical and methodological reflections on research into translation processes. In F. Alves (Ed.), *Triangulating Translation: Perspectives in Process Oriented Research* (pp. 25–42). Amsterdam: John Benjamins.

Hansen, G. (2006). *Erfolgreich Übersetzen. Entdecken und Beheben von Störquellen*. Tübingen: Narr Francke Attempto.

Hansen, G. (2010). Some thoughts about the evaluation of translation products in empirical translation process research. In I. M. Mees, F. Alves, & S. Göpferich (Eds.), *Methodology, technology and innovation in translation process research : a tribute to Arnt Lykke Jakobsen* (pp. 389–402). Frederiksberg: Samfundslitteratur.

Haug, T., & Audeoud, M. (2013). 'Directionality Effect' beim Gebärdensprachdolmetschen – Welche Dolmetschrichtung präferieren Gebärdensprachdolmetscher in der Schweiz? Teil 2. *Das Zeichen*, *27*(94), 306–316.

Heidrich, F. (2016). *Kommunikationsoptimierung im Fachübersetzungsprozess*. Berlin: Frank & Timme.

Hekkanen, R. (2007). The role of ethics in translation and in translation studies research. *Across Languages and Cultures*, *8*(2), 231–247. https://akjournals.com/view/journals/084/8/2/article-p231.xml

Heredia, R. R., & Brown, J. M. (2004). Bilingual memory. *The Handbook of Bilingualism*, 225–249.

Hirci, N. (2009). The efficacy of using state of-the-art translation resources in translating into a non-mother tongue. *New Voices in Translation Studies*, *5*, i–iii.

Hirci, N. (2012). Electronic reference resources for translators: Implications for productivity and translation quality. *Interpreter and Translator Trainer*, *6*, 219–236. https://doi.org/10.1080/13556509.2012.10798837

House, J. (1997). *Translation quality assessment: A model revisited*. Tübingen: Gunter Narr.

House, J. (2015). *Translation quality assessment: past and present*. London: Routledge. https://doi.org/10.1057/9781137025487_13

Hung, E. (2005). Collaborative translation – or what can we learn from the Chinese translation tradition? In W. Pan (Ed.), *Translation and Contrastive Studies* (pp. 21–45). Shanghai: Waiyu Jiaoyu.

Hunziker Heeb, A. (2012). The problem-solving processes of experienced and non-experienced translators. In S. Kersten, C. Ludwig, D. Meer, & B. Rüschoff (Eds.), *Language learning and language use – applied linguistics approaches. Papers selected from the Junior Research Meeting – Essen 2011* (pp. 177–186). Duisburg: UVRR.

Hunziker Heeb, A. (2015). Language processing in professional L2 translation: indication from research behaviour. Presented at the Poznań Linguistic Meeting, 17–19 September 2015, Adam Mickiewicz University, Poznań.

Hunziker Heeb, A. (2016). Professional translators' self-concepts and directionality: indications from translation process research. *The Journal of Specialised Translation*, (25), 74–88. https://doi.org/10.21256/zhaw-2049

Hurtado Albir, A. (Ed.). (2017). *Researching translation competence by PACTE group*. Amsterdam: John Benjamins. https://doi.org/10.1075/btl.127

Hvelplund, K. T. (2011). *Allocation of cognitive resources in translation. An eye-tracking and key-logging study*. Doctoral dissertation. Copenhagen Business School.

Hvelplund, K. T. (2014). Eye tracking and the translation process: Reflections on the analysis and interpretation of eye-tracking data. *MonTI Monografías de Traducción e Interpretación, Special Issue-Minding Translation*, 201–223. http://dx.doi.org/10.6035/MonTI.2014.ne1.6

Hvelplund, K. T. (2017a). Four fundamental types of reading during translation. In A. L. Jakobsen & B. Mesa-Lao (Eds.), *Translation in Transition: Between Cognition, Computing and Technology* (pp. 55–77). Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/btl.133.02hve

Hvelplund, K. T. (2017b). Translators' use of digital resources during translation. *HERMES - Journal of Language and Communication in Business*, (56), 71–87. https://doi.org/10.7146/hjlcb.v0i56.97205

Hvelplund, K. T., & Dragsted, B. (2017). Genre familiarity and translation processing: Differences and similarities between literary and LSP translators. In I. Lacruz & R. Jääskeläinen (Eds.), *Innovation and expansion in translation process research* (Vol. XVIII, pp. 55–76). Amsterdam/Philadelphia: John Benjamins.

IAPTI's Ethics Committee. (2015). *Translation into a non-native language*. IAPTI – International Association of Professional Translators and Interpreters. Retrieved from https://www.iapti.org/files/surveys/2/IAPTI_non-native_report.pdf

ISO. (2015). ISO 17100: Translation services — Requirements for translation services.

ISO. (2017). ISO 18587: Translation services — Post-editing of machine translation output — Requirements.

Jääskeläinen, R. (1999). *Tapping the process. An explorative study of the cognitive and affective factors involved in translation*. Joensuu: Joensuun Yliopisto.

Jääskeläinen, R. (2010). Are all professionals experts? Definitions of expertise and reinterpretation of research evidence in process studies. In G. M. Shreve & E. Angelone (Eds.), *Translation and cognition* (pp. 213–227). Amsterdam: John Benjamins. https://doi.org/10.1075/ata.xv.12jaa

Jääskeläinen, R. (2016). Quality and translation process research. In R. Muñoz Martín (Ed.), *Reembedding translation process research* (Vol. 128, pp. 89–106). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.128.05jaa

Jääskeläinen, R., Kujamäki, P., & Mäkisalo, J. (2011). Towards professionalism – or against it? Dealing with the changing world in translation research and translator education. *Across Languages and Cultures*, *12*(2), 143–156. https://akjournals.com/view/journals/084/12/2/article-p143.xml

Jakobsen, A. L. (2002). Translation drafting by professional translators and by translation students. In G. Hansen (Ed.), *Empirical translation studies: Process and product* (Vol. 27, pp. 191–204). Copenhagen: Samfundslitteratur.

Jakobsen, A. L. (2014). The development and current state of translation process research. In E. Brems, R. Meylaerts, & L. van Doorslaer (Eds.), *The known unknowns of translation studies* (pp. 65–88). John Benjamins.

Jakobsen, A. L., & Jensen, K. T. H. (2008). Eye movement behaviour across four different types of reading task. In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), *Looking at eyes. Eye-tracking studies of reading and translation processing* (pp. 103–124). Frederiksberg: Samfundslitteratur.

Jamoussi, R. (2015). Exporting cultural goods through the medium of translation in the Arab world: the (not so) strange case of L2 translation. *The Translator*, *21*(2), 173–188. https://doi.org/10.1080/13556509.2015.1069044

Kearns, J. (2007). Translate and be damned! Inverse translation and professional recognition. In R. Sokoloski, H. Duda, K. Klimkowski, & J. Klimek (Eds.), *Warsztaty translatorskie / Workshop on translation* (Vol. 4, pp. 171–191). Lublin/Ottawa: Towarzystwo Naukowe Katolickiego Uniwersytetu Lubelskiego Jana Pawła II/Slavic Research Group at the University of Ottawa.

Kearns, J. (2012). Curriculum ideologies in translator and interpreter training. In S. Hubscher-Davidson & M. Borodo (Eds.), *Global trends in translator and interpreter training* (pp. 11–29). London: Continuum.

Keller, L. (2018). *Targeted language selection in the multilingual mind: Language control and language-variety selection in simultaneous interpreting*. University of Geneva, Geneva. https://doi.org/10.13097/archive-ouverte/unige:107361

Kelly, D., Martin, A., Nobs, M.-L., Sánchez, D., & Way, C. (Eds.). (2003). *La direccionalidad en traducción e interpretación: perspectivas teóricas, profesionales y didácticas*. Granada: Atrio.

Kelly, D., Nobs, M.-L., Sánchez, D., & Way, C. (2003). La traducción A-B en el mercado profesional. In D. Kelly, A. Martin, M.-L. Nobs, D. Sánchez, & C. Way (Eds.), *La direccionalidad en traducción e interpretación: perspectivas teóricas, profesionales y didácticas* (pp. 43–64). Granada: Atrio.

Keyton, J. (2014). *Communication research: Asking questions, finding answers* (4th ed.). New York: McGraw-Hill.

Kiraly, D. (1995). *Pathways to translation: Pedagogy and process*. Kent: Kent State University Press. https://doi.org/10.7202/002159ar

Kiraly, D. (1997). Think-aloud protocols and the construction of a professional translator self-concept. In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (Vol. 3, pp. 137–160). Thousand Oaks: Sage.

Kiraly, D. (2000). Translation into a non-mother tongue: From collaboration to competence. In M. Grosman, M. Kadrić, I. Kovačič, & M. Snell-Hornby (Eds.), *Translation into non-mother tongues In professional practice and training* (2nd ed., Vol. 8, pp. 117–123). Tübingen: Stauffenburg.

Kiraly, D. (2005). Project-based learning: A case for situated translation. *Meta*, *50*(4), 1098–1111. https://doi.org/10.7202/012063ar

Kirsh, D. (2009). Problem solving and situated cognition. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 264–306). Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511816826

Klimant, P. (2016). Translation evaluation upside down: Phenomena instead of errors. In C. Martín de León & V. González-Ruiz (Eds.), *From the Lab to the Classroom and Back Again* (pp. 181–220). Peter Lang. https://doi.org/10.3726/978-3-0353-0806-8

Krings, H. P. (1986). Sekundäre Indikatoren. In H. P. Krings (Ed.), *Was in den Köpfen von Übersetzern vorgeht* (pp. 130–143). Tübingen: Gunter Narr.

Krings, H. P. (2001). *Repairing texts: Empirical investigations of machine translation post-editing processes*. (G. S. Koby, G. M. Shreve, K. Mischerikow, & S. Litzer, Trans.). Kent: Kent State University Press. https://doi.org/10.7202/008026ar

Krings, H. P. (2005). Wege ins Labyrinth – Fragestellungen und Methoden der Übersetzungsprozessforschung im Überblick. *Meta*, *50*(2), 342–358.

Kruger, H. (2016). What's happening when nothing's happening? Combining eyetracking and keylogging to explore cognitive processing during pauses in translation production. *Across Languages and Cultures*, *17*(1), 25–52. https://doi.org/10.1556/084.2016.17.1.2

Kruger, H., & Kruger, J.-L. (2017). Cognition and Reception. In J. W. Schwieter & A. Ferreira (Eds.), *The Handbook of Translation and Cognition* (pp. 71–89). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119241485.ch4

Krüger, R. (2015). Fachübersetzen aus kognitionstranslatologischer Perspektive. Das Kölner Modell des situierten Fachübersetzers. *TransKom Zeitschrift Für Translationswissenschaft Und Fachkommunikation*, *8*(2), 273–313.

Kuckartz, U. (2014). *Mixed methods*. Wiesbaden: Springer. https://doi.org/10.1007/978-3-531-93267-5

Kumpulainen, M. (2015). On the operationalisation of 'pauses' in translation process research. *Translation & Interpreting*, *7*(1), 47–58. http://www.trans-int.org/index.php/transint/article/viewFile/367/183

Kumpulainen, M. (2016). *Learning translation: an empirical study into the acquisition of interlingual text production skills*. Doctoral dissertation. University of Eastern Finland.

Kuznik, A. (2017a). Characteristics of the sample. In A. Hurtado Albir (Ed.), *Researching translation competence by PACTE group* (pp. 97–105). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.127.06kuz

Kuznik, A. (2017b). Use of instrumental resources. In A. Hurtado Albir (Ed.), *Researching translation competence by PACTE group* (pp. 219–242). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.127.15kuz

Kuznik, A., & Olalla-Soler, C. (2018). Results of PACTE group's experimental research on translation competence acquisition. The acquisition of the instrumental sub-competence. *Across Languages and Cultures*, *19*, 19–51. https://doi.org/10.1556/084.2018.19.1.2

Lacruz, I. (2017). Cognitive effort in translation, editing, and post-editing. In J. W. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 386–401). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119241485.ch21

Läubli, S., Sennrich, R., & Volk, M. (2018). Has machine translation achieved human parity? A case for document-level evaluation. *ArXiv:1808.07048 [Cs]*. Retrieved from http://arxiv.org/abs/1808.07048

Lavault-Olléon, É. (Ed.). (2016). *Approches ergonomiques des pratiques professionnelles et des formations des traducteurs. ILCEA [special issue]. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*. Grenoble: Ellug. Retrieved from https://journals.openedition.org/ilcea/3834

Leijten, M., Van Waes, L., & Van Horenbeeck, E. (2015). Analyzing writing process data. A linguistic perspective. In G. Cislaru (Ed.), *Writing(s) at the crossroads: the process-product interface* (pp. 277–302). Amsterdam: John Benjamins. https://doi.org/10.1075/z.194.14lei

Livbjerg, I., & Mees, I.M. (2003). Patterns of dictionary use in non-domain-specific translation. In F. Alves (Ed.), *Triangulating Translation: Perspectives in Process Oriented Research* (pp. 123–136). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.45.11liv

Lorenzo, M.P. (1999). La seguridad del traductor profesional en la traduccion a una lengua extranjera. In G. Hansen (Ed.), *Probing the process in translation: methods and results* (pp. 121–134). Copenhagen: Samfundslitteratur.

Lorenzo, M.P. (2002a). Competencia revisora y traducción inversa. *Cadernos de Tradução*, *2*(10), 133–166. https://periodicos.ufsc.br/index.php/traducao/article/viewFile/6148/5706

Lorenzo, M.P. (2002b). ¿Es posible la traducción inversa? Resultados de un experimento sobre traducción profesional a una lengua extranjera. In G. Hansen (Ed.), *Empirical Translation Studies – Process and Product* (pp. 85–124). Copenhagen: Samfundslitteratur.

Lorenzo, M.P. (2003). La traducción a una lengua extranjera: uno de los muchos desafíos a la competencia traductora. In D. Kelly, A. Martin, M.-L. Nobs, D. Sánchez, & C. Way (Eds.), *La direccionalidad en traducción e interpretación: perspecitvas teóricas, profesionales y didácticas*. Granada: Atrio.

Lörscher, W. (1991). Investigating the translation process. In *Translation performance, translation process, and translation strategies. A psycholinguistic investigation*. Tübingen: Gunter Narr. https://doi.org/10.7202/003517ar

Macizo, P., & Bajo, M.T. (2006). Reading for repetition and reading for translation: do they involve the same processes? *Cognition*, *99*(1), 1–34. https://doi.org/10.1016/j.cognition.2004.09.012

Marín García, A. (2017). *Theoretical hedging: The scope of knowledge in Translation Process Research*. Doctoral dissertation. Kent State University.

Martín de León, C. (2008). Skopos and beyond. A critical study of functionalism. *Target*, *20*(1), 1–28. https://doi.org/10.1075/target.20.1.02mar

Martín de León, C. (2017). Mental representations. In J. W. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 106–126). Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9781119241485.ch6

Martin, S. (2014). Measuring cognitive load and cognition: metrics for technology-enhanced learning. *Educational Research and Evaluation*, *20*(7–8), 592–621. https://doi.org/10.1080/13803611.2014.997140

Massey, G., & Ehrensberger-Dow, M. (2012). Evaluating the process: Implications for curriculum development. In L. N. Zybatow, A. Petrova, & M. Ustaszewski (Eds.), *Translationwissenschaft interdisziplinär: Fragen der Theorie und Didaktik / Translation studies: Interdisciplinary issues in theory and didactics* (pp. 95–100). Frankfurt am Main: Peter Lang.

Massey, G., & Ehrensberger-Dow, M. (2014). Looking beyond text. The usefulness of translation process data. In D. Knorr, C. Heine, & J. Engberg (Eds.), *Methods in writing process research* (Vol. 13, pp. 81–98). Frankfurt am Main: Peter Lang.

McAlester, G. (1992). Teching translation into a foreign language – status, scope and aims. In C. Dollerup & A. Loddegaard (Eds.), *Teaching translation and interpreting. Training, talent and experience. Papers from the First Language International Conference. Elsinore, Denmark, 31 May – 2 June 1991* (pp. 291–297). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/z.56.44mca

McAlester, G. (2000a). The evaluation of translation into a foreign language. In C. Schäffner & B. Adab (Eds.), *Developing translation competence* (pp. 229–241). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.38.21mca

McAlester, G. (2000b). The time factor: A practical evaluation criterion. In M. Grosman, M. Kadrić, I. Kovačič, & M. Snell-Hornby (Eds.), *Translation into non-mother tongues: in professional practice and training* (2nd ed., pp. 133–139). Tübingen: Stauffenburg.

Mellinger, C. D., & Hanson, T. A. (2016). *Quantitative research methods in translation and interpreting studies*. Milton Park, Abingdon, Oxon; New York, NY: Routledge.

Moorkens, J. (2018). Eye tracking as a measure of cognitive effort for post-editing of machine translation. In C. Walker & F. M. Federici (Eds.), *Benjamins Translation Library* (Vol. 143, pp. 55–70). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.143.04moo

Mossop, B. (2014). *Revising and editing for translators* (3rd ed.). London: Routledge. https://doi.org/10.4324/9781315767130

Muñoz Martín, R. (2010). On paradigms and cognitive translatology. In G. M. Shreve & E. Angelone (Eds.), *Translation and Cognition* (pp. 169–187). Amsterdam: John Benjamins. https://doi.org/10.1075/ata.xv.10mun

Muñoz Martín, R. (2012a). Just a matter of scope. Mental load in translation process research. *Translation Spaces*, *1*(1), 169–188. https://doi.org/10.1075/ts.1.08mun

Muñoz Martín, R. (2012b). Standardizing translation process research methods and reports. In I. García Izquierdo & E. Monzó (Eds.), *Iberian Studies on Translation and Interpreting* (Vol. 11, pp. 11–22). Bern: Peter Lang.

Muñoz Martín, R. (2014). Situating translation expertise: A review with a sketch of a construct. In J.W. Schwieter & A. Ferreira (Eds.), *The development of translation competence: theories and methodologies from psycholinguistics and cognitive science* (pp. 2–56). Newcastle upon Tyne: Cambridge Scholars Publishing.

Muñoz Martín, R. (2016). Processes of what models? On the cognitive indivisibility of translation acts and events. *Translation Spaces*, *5*(1), 145–161. https://doi.org/10.1075/ts.5.1.08mun

Muñoz Martín, R. (2017). Looking toward the future of cognitive translation studies. In J. W. Schwieter & A. Ferreira (Eds.), *The Handbook of Translation and Cognition* (pp. 555–572). Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119241485.ch30

Muñoz Martín, R., & Cardona Guerra, J. M. (2018). Translating in fits and starts: pause thresholds and roles in the research of translation processes. *Perspectives*, 1–27. https://doi.org/10.1080/0907676X.2018.1531897

Muñoz Martín, R., & Conde Ruano, J. T. (2007). Effects of serial translation evaluation. In P. A. Schmitt & H. E. Jüngst (Eds.), *Translationsqualität* (Vol. 5, pp. 428–444).

Muñoz Martín, R., & Martín de León, C. (2018). Fascinatin' rhythm – and pauses in translators' cognitive processes. *HERMES – Journal of Language and Communication in Business*, (57), 29–47. https://doi.org/10.7146/hjlcb.v0i57.106192

Murray, H. (2003). Swiss English teachers and Euro-English: Attitudes to a non-native variety. *Bulletin Suisse de Linguistique Appliquée (VALS/ASLA)*, 147–165.

Neunzig, W. (2011). Empirical studies in translation: Methodological and epistemological questions. *TTR : Traduction, Terminologie, Rédaction*, *24*(2), 15–39. https://doi.org/10.7202/1013393ar

Neunzig, W., & Tanqueiro, H. (2007). Risikominimierung beim Übersetzen in die Fremdsprache: ein Thesenpapier aus spanischer Sicht. In G. Wotjak (Ed.), *Quo vadis Translatologie? Ein halbes Jahrhundert universitäre Ausbildung von Dolmetschern und Übersetzern in Leipzig. Rückschau, Zwischenbilanz und Perspektive aus der Außensicht.* (pp. 279–293). Berlin: Frank & Timme.

Nida, E. (1964). *Toward a Science of Translating*. Leiden: E. J. Brill.

Nobs, M.-L. (2003). *Expectativas y evaluación en la traducción de folletos turísticos: estudio empírico con usuarios reales*. Doctoral dissertation. Universidad de Granada.

Nord, B. (2002). *Hilfsmittel beim Übersetzen. Eine empirische Studie zum Rechercheverhalten professioneller Übersetzer*. Frankfurt am Main: Peter Lang.

Nord, C. (1991). *Text analysis in translation*. Amsterdam/Atlanta: Rodopi.

Nørgård, R. T., Toft-Nielsen, C., & Whitton, N. (2017). Playful learning in higher education: developing a signature pedagogy. *International Journal of Play*, *6*(3), 272–282. https://doi.org/10.1080/21594937.2017.1382997

O'Brien, S. (2006a). Eye-tracking and translation memory matches. *Perspectives: Studies in Translatology*, *14*(3), 185–205. https://doi.org/10.1080/09076760708669037

O'Brien, S. (2006b). Pauses as indicators of cognitive effort in post-editing machine translation output. *Across Languages and Cultures*, *7*(1), 1–21. http://dx.doi.ord/10.1556/ACROSS.7.2006.1

O'Brien, S. (2010). Eye tracking in translation process research: methodological challenges and solutions. In I. M. Mees, F. Alves, & S. Göpferich (Eds.), *Methodology, technology and innovation in translation process research: a tribute to Arnt Lykke Jakobsen* (pp. 251–266). Frederiksberg: Samfundslitteratur.

O'Brien, S. (2013). The borrowers: Researching the cognitive aspects of translation. *Target*, *25*(1), 5–17. https://doi.org/10.1075/target.25.1.02obr

Olalla-Soler, C. (2018). Using electronic information resources to solve cultural translation problems: Differences between students and professional translators. *Journal of Documentation*, *74*(6), 1293–1317. https://doi.org/10.1108/JD-02-2018-0033

Olive, T., & Cislaru, G. (2015). Linguistic forms at the process-product interface. In G. Cislaru (Ed.), *Writing(s) at the crossroads: the process-product interface* (pp. 99–124). Amsterdam: John Benjamins. https://doi.org/10.1075/z.194.06oli

Opdenhoff, J.-H. (2011). *Estudio sobre la direccionalidad en interpretación de conferencias: de las teorías a la práctica profesional*. Doctoral dissertation. Universidad de Granada.

Opdenhoff, J.-H. (2013). Interpreting quality in the light of directionality – A study on the interpreter's perspective. In O. García Becerra, E. M. Pradas Macías, & R. Barranco-Droege (Eds.), *Quality in interpreting – Widening the scope* (Vol. 1, pp. 201–220). Granada: Comares.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8

PACTE. (2003). Building a translation competence model. In F. Alves (Ed.), *Triangulating Translation. Perspectives in process oriented research* (pp. 43–66). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.45.06pac

PACTE. (2005). Investigating translation competence: Conceptual and methodological issues. *Meta*, *50*(2), 609–619. https://doi.org/10.7202/011004ar

PACTE. (2009). Results of the validation of the PACTE translation competence model: Acceptability and decision making. *Across Languages and Cultures*, *10*(2), 207–230.

PACTE. (2011). Results of the validation of the PACTE translation competence model. Translation problems and translation competence. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of process research* (pp. 317–343). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/btl.94.22pac

PACTE. (2017a). Conclusions. Defining features of translation competence. In A. Hurtado Albir (Ed.), *Researching translation competence by PACTE group* (pp. 281–303). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.127.18pac

PACTE. (2017b). Experiment design. In A. Hurtado Albir (Ed.), *Researching translation competence by PACTE group* (pp. 81–96). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.127.18pac

Pavlović, N. (2007a). *Directionality in collaborative translation processes. A study of novice translators*. Doctoral dissertation. Universitat Rovira i Virgili.

Pavlović, N. (2007b). Directionality in translation and interpreting practice. Report on a questionnaire survey in Croatia. *Forum*, *5*(2), 79–99. https://doi.org/10.1075/forum.5.2.05pav

Pavlović, N. (2010). What were they thinking?! Students' decision making in L1 and L2 translation processes. *HERMES- Journal of Language and Communication Studies*, (44), 63–87. https://doi.org/10.7146/hjlcb.v23i44.97267

Pavlović, N., & Jensen, K. T. H. (2009). Eye tracking translation directionality. In A. Pym & A. Perekrestenko (Eds.), *Translation research projects 2* (pp. 93–109). Tarragona: Intercultural Studies Group. Retrieved from http://isg.urv.es/publicity/isg/publications/trp_2_2009/index.htm

Pedersen, D. (2017). Managing transcreation projects: An ethnographic study. *Translation Spaces, Translation Spaces. A Multidisciplinary, Multimedia, and Multilingual Journal of Translation*, *6*(1), 44–61. https://doi.org/10.1075/ts.6.1.03ped

Phakiti, A., & Paltridge, B. (2015). Approaches and methods in applied linguistics research. In B. Paltridge & A. Phakiti (Eds.), *Research methods in applied linguistics: a practical resource* (pp. 5–25). London: Bloomsbury.

Piolat, A., Roussey, J.-Y., Olive, T., & Amada, M. (2004). Processing time and cognitive effort in revision: Effects of error type and of working memory capacity. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes* (pp. 21–38). New York: Springer. https://doi.org/10.1007/978-94-007-1048-1_3

Pokorn, N. K. (2000). Translation into a non-mother tongue in translation theory: deconstruction of the traditional. In A. Chesterman, N. Gallardo San Salvador, & Y. Gambier (Eds.), *Translation in context* (pp. 61–72). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.39.09pok

Pokorn, N. K. (2005). *Challenging the traditional axioms: translation into a non-mother tongue* (Vol. 62, 3). Amsterdam: John Benjamins. https://doi.org/https://doi.org/10.1075/btl.62

Pokorn, N. K. (2008). Translation and TS research in a culture using a language of limited diffusion: The case of Slovenia. *The Journal of Specialised Translation*, *10*, 2–9.

Pokorn, N. K. (2009). Natives or non-natives? That is the question ... Teachers of translation into language B. *The Interpreter and Translator Trainer*, *3*(2), 189–208. https://doi.org/10.1080/1750399X.2009.10798788

Prassl, F. (2011). Übersetzerisches Rechercheverhalten von Profis und Novizen auf dem Prüfstand. *Trans-Kom*, *4*(1), 23–48.

Pym, A. (2011). Translation research terms: a tentative glossary for moments of perplexity and dispute. In A. Pym (Ed.), *Translation Research Projects 3* (pp. 75–110). Tarragona: Intercultural Studies Group.

Pym, A., Grin, F., Sfreddo, C., & Chan, A. L. J. (2013). *The status of the translation profession in the European Union* (Vol. 7). https://doi.org/10.2782/63429

Rayner, K. (1997). Understanding eye movements in reading. *Scientific Studies of Reading*, *1*(4), 317–339. https://doi.org/10.1207/s1532799xssr0104_2

Risku, H. (1998). *Translatorische Kompetenz: kognitive Grundlagen des Übersetzens als Expertentätigkeit* (Vol. 5). Tübingen: Stauffenburg.

Risku, H. (2002). Translations- und Kognitionswissenschaftliche Paradigmen: der Mensch im Mittelpunkt. In L. Zybatow (Ed.), *Translationswissenschaft im interdisziplinären Dialog. Innsbrucker Ringvorlesungen zur Translationswissenschaft III* (Vol. 2329, pp. 55–69). Tübingen: Francke.

Risku, H. (2016). *Translationsmanagement: interkulturelle Fachkommunikation im Informationszeitalter* (3rd rev. ed., Vol. 1). Tübingen: Narr.

Risku, H., Milosevic, J., & Pein-Weber, C. (2016). Writing vs. translating. In R. Muñoz Martín (Ed.), *Reembedding Translation Process Research* (Vol. 128, pp. 47–68). John Benjamins. https://doi.org/10.1075/btl.128

Risku, H., Rogl, R., & Pein-Weber, C. (2016). Mutual dependencies: centrality in translation networks. *The Journal of Specialised Translation*, (25), 1–22.

Robson, C. (2002). *Real World Research*. Victoria: Blackwell Publishing.

Rodríguez-Inés, P. (2014). Using corpora for awareness-raising purposes in translation, especially into a foreign language (Spanish-English). *Perspectives: Studies in Translatology*, *22*(2), 222–241. https://doi.org/10.1080/0907676X.2012.685239

Rodríguez-Inés, P., & Fox, O. (2018). Reconciling institutional and professional requirements in the specialised inverse translation class – A case study. *Meta*, *63*(1), 47–71. https://doi.org/10.7202/1050514ar

Ruiz, C., Paredes, N., Macizo, P., & Bajo, M. T. (2008). Activation of lexical and syntactic target language properties in translation. *Acta Psychologica*, *128*(3), 490–500. https://doi.org/10.1016/j.actpsy.2007.08.004

Sakamoto, A. (2017). Professional translators' theorising patterns in comparison with classroom discourse on translation: The case of Japanese/English translators in the UK. *Meta*, *62*(2), 271–288. https://doi.org/10.7202/1041024ar

Saldanha, G., & O'Brien, S. (2013). *Research methodologies in translation studies*. London: Routledge. https://doi.org/https://doi.org/10.4324/9781315760100

Schaeffer, M., & Carl, M. (2014). Measuring the cognitive effort of literal translation processes. In U. Germann, M. Carl, P. Koehn, G. Sanchis-Trilles, F. Casacuberta, R. Hill, & S. O'Brien (Eds.), *Proceedings of the Workshop on Humans and Computer-assisted Translation (HaCaT)* (pp. 29–37). Stroudsburg, PA: Association for Computational Linguistics. https://doi.org/10.3115/v1/w14-0306

Schäffner, C. (2000). Running before walking? Designing a translation programme at undergraduate level. In C. Schäffner & B. Adab (Eds.), *Developing translation competence* (Vol. 38, pp. 143–156). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.38.14sch

Schmitt, P. A., Gerstmeyer, L., & Müller, S. (2016). Übersetzer und Dolmetscher: eine internationale Umfrage zur Berufspraxis. Berlin: BDÜ Fachverlag.

Schrijver, I. (2014). *The translator as a text producer: the effects of writing training on transediting and translation performance*. Doctoral dissertation. Universiteit Antwerpen.

Schubert, K. (2007). *Wissen, Sprache, Medium, Arbeit: ein integratives Modell der ein- und mehrsprachigen Fachkommunikation*. Tübingen: Narr.

Schubert, K. (2009). Positioning translation in technical communication studies. *The Journal of Specialised Translation*, *11*, 17–30.

Schubert, K. (2013). Die fachkommunikative Arbeitsprozesskette. In M. Krein-Kühle, U. Wienen, & R. Krüger (Eds.), *Kölner Konferenz zur Fachtextübersetzung (2010)* (pp. 151–162). Frankfurt am Main: Peter Lang. https://doi.org/10.3726/978-3-653-02775-4

Schwieter, J. W., & Ferreira, A. (Eds.). (2017). *The handbook of translation and cognition*. London: Wiley-Blackwell. https://doi.org/10.1002/9781119241485

Šebesta, D. (2013). *Native speaker status in the translation services market*. Lambert Academic Publishing.

Seeber, K. G. (2011). Cognitive load in simultaneous interpreting. Existing theories – new models. *Interpreting*, *13*(2), 176–204. https://doi.org/10.1075/intp.13.2.02see

Sela-Sheffy, R. (2016). Profession, identity, and status. In C. V. Angelelli & B. J. Baer (Eds.), *Researching translation and interpreting* (Vol. 1, pp. 131–145). New York/Oxon: Routledge.

Selvi, A. F. (2014). Myths and misconceptions about nonnative English speakers in the TESOL (NNEST) movement. *TESOL Journal*, *5*(3), 573–611. https://doi.org/10.1002/tesj.158

Shreve, G. M. (2006). The deliberate practice: Translation and expertise. *Journal of Translation Studies*, *9*(1), 27–42.

Shreve, G. M. (2009). Recipient-orientation and metacognition in the translation process. In R. Dimitriu & M. Shlesinger (Eds.), *Translators and their readers: in homage to Eugène A. Nida* (pp. 255–270). Bruxelles: Éditions du Hazard.

Si, L., & Callan, J. (2001). A statistical model for scientific readability. In *CIKM '01 Proceedings of the Tenth International Conference on Information and Knowledge Management* (pp. 574–576). New York: ACM Press. https://doi.org/10.1145/502692.502695

Simons, G. F., & Fennig, C. D. (2018). Ethnologue: Languages of the world. Retrieved 6 February 2019, from http://www.ethnologue.com/language/eng

Sjørup, A. C. (2013). *Cognitive effort in metaphor translation: an eye-tracking and key-logging study*. Doctoral dissertation. Copenhagen Business School.

Suojanen, T., Koskinen, K., & Tuominen, T. (2015). *User-centered translation*. London and New York: Routledge. Retrieved from http://www.tandfebooks.com/action/showBook?doi=10.4324/9781315753508

TechSmith. (n. d.). Camtasia. Die beste Komplettlösung für Bildschirmaufnahme und Videobearbeitung. Retrieved 7 July 2013, from http://www.techsmith.de/camtasia.asp

Teixeira, C. S. C. (2014). *The impact of metadata on translator performance: How translators work with translation memories and machine translation*. Doctoral dissertation. Universitat Rovira i Virgili. https://doi.org/10.13140/RG.2.1.2190.2887

Tirkkonen-Condit, S. (2000). Challenges and priorities in process research. In S. Tirkkonen-Condit & R. Jääskeläinen (Eds.), *Tapping and mapping the processes of translation and interpreting: outlooks on empirical research* (Vol. 37, pp. vii–ix). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.37.01tir

Tobii. (2019). The world leader in eye tracking. Retrieved 12 March 2019, from http://www.tobii.com

Tymoczko, M. (2006). Reconceptualizing translation theory: integrating non-Western thought about translation. In T. Hermans (Ed.), *Translating Others* (Vol. 1, pp. 13–32). Manchester: St. Jerome.

Tymoczko, M. (2018). The history of internationalization in translation studies and its impact on translation theory. In L. D'Hulst & Y. Gambier (Eds.), *A history of modern translation knowledge: sources, concepts, effects* (pp. 153–168). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/btl.142.21tym

Van Waes, L., & Leijten, M. (2006). Logging writing processes with Inputlog. In L. Van Waes, M. Leijten, & C. Neuwirth (Eds.), *Writing and digital media* (Vol. 17, pp. 158–166). Oxford: Elsevier.

Vieira, L. N. (2016a). *Cognitive effort in post-editing of machine translation: Evidence from eye movements, subjective ratings, and think-aloud protocols*. Doctoral dissertation. Newcastle University.

Vieira, L. N. (2016b). How do measures of cognitive effort relate to each other? A multivariate analysis of post-editing process data. *Machine Translation*, *30*(1), 41–62. https://doi.org/10.1007/s10590-016-9188-5

Vieira, L. N. (2017). From process to product: Links between post-editing effort and post-edited quality. In A. L. Jakobsen & B. Mesa-Lao (Eds.), *Benjamins Translation Library* (Vol. 133, pp. 162–186). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.133.06vie

Wang, B. (2011). Translation practices and the issue of directionality in China. *Meta*, *56*(4), 896. https://doi.org/10.7202/1011259ar

Wang, J., & Napier, J. (2015). Directionality in signed language interpreting. *Meta*, *60*(3), 518–541. https://doi.org/10.7202/1036141ar

Whyatt, B. (2018). Testing indicators of translation expertise in an intralingual task. *HERMES – Journal of Language and Communication in Business*, (57), 63–78. https://doi.org/10.7146/hjlcb.v0i57.106194

Whyatt, B., Kajzer-Wietrzny, M., & Stachowiak, K. (2017). Intralingual and interlingual translation. Designing a comparative study of decision-making processes in paraphrase and translation. In A. L. Jakobsen & B. Mesa-Lao (Eds.), *Translation in transition: Between cognition, computing and technology* (pp. 135–158). Amsterdam: John Benjamins. https://doi.org/10.1075/btl.133.05why

Whyatt, B., & Kościuczuk, T. (2013). Translation into a non-native language: The double life of the native-speakership axiom. *MTm. Translation in an Age of Austerity*, (5), 60–79.

Whyatt, B., Kościuczuk, T., & Turski, M. (2017). Incorporating proof-readers to understand the effects of directionality on the translation process. Poster session presented at the Translation & Interpreting in Transition 3, 13–14 July 2017, Ghent University.

Wimmer, S. (2011). *El proceso de la traducción especializada inversa: modelo, validación empírica y aplicación didáctica*. Doctoral dissertation. Universitat Autònoma de Barcelona.

XLSTAT. (2017). Retrieved 17 May 2018, from https://www.xlstat.com.

11 Appendices

Table of contents

Appendix A: Main study on translation processes 220
A1 Source texts and translation briefs 220
A2 Semi-structured interviews 221
A3 Introduction of participant and workplace set-up ... 225
A4 Using eye tracking as a data collection method 226

Appendix B: Substudy on product quality 228

Appendix C: Results for translators' processes 233
C1 Data preparation and data analysis of main study 233
C2 Global process measures 237
C3 Total numbers of process activities 239
C4 Writing in the translation process 240
C5 Revising in the translation process 243
C6 Information seeking in the translation process 247
C7 Pausing in the translation process 252

Appendix D: Results for translator's products 255
D1 Target texts BiDir 1-6 L1-L2 (English) 255
D2 Target texts BiDir 1-6 L2-L1 (German) 257
D3 Target texts UniEnglish 1-6 L2- L1 259
D4 Target texts UniGerman 1-6 L2- L1 261
D5 Data preparation and data analysis of substudy 264
D6 Global measures of translation products 266
D7 Product quality of English target texts 267

Appendix E: Results for translators' effort 271
E1 Data preparation and data analysis of translator effort 272
E2 Character production effort 275
E3 Revision effort 276
E4 Information search effort 277
E5 Coordination effort 278
E6 Mean fixation duration 279

Appendix F: Results for relations between characteristics of professional translation 280

Appendix A: Main study on translation processes

This appendix contains information on the stimuli, the interview questionnaires and the procedure of the main study (see Section 4.2 of the chapter *Methodology*). It also contains participant data and an overview of the data collected.

A1 Source texts and translation briefs

German source text and translation brief

Übersetzungsauftrag: Zu übersetzen ist ein Abschnitt aus einem Artikel, der im April 2009 in der Tageszeitung *Neue Zürcher Zeitung* erschienen ist. Der Text soll für eine ähnliche Tageszeitung der Zielkultur übersetzt werden.

Strandungen von Walen
Ein Hang zum Selbstmord dürfte dem Phänomen nicht zugrunde liegen. Vielmehr sind es wohl meist mehrere und oft von Fall zu Fall verschiedene Faktoren, die Strandungen lebender Wale verursachen oder begünstigen. Die am besten untersuchten Strandungen sind die von Schnabelwalen, für die ein Zusammenhang mit dem Einsatz bestimmter Sonartypen vermutet wird. Nach solchen Sonareinsätzen beobachtete man mehrfach ein für die Gattung ungewöhnliches Strandungsmuster: Viele Schnabelwale strandeten innert weniger Stunden, über viele Kilometer Küstenlinie verstreut. Bei manchen von ihnen stellten die Forscher Verletzungen der Hörorgane fest, die auf einen Verlust der Navigationsfähigkeit schliessen lassen.

(Anzahl Wörter: 96)

English source text and translation brief

Übersetzungsauftrag: Zu übersetzen ist ein Abschnitt aus einem Artikel von Mark Townsend, der im August 2004 in der Onlinezeitung *The Observer* erschienen ist. Der Text soll für eine ähnliche deutschsprachige Tageszeitung übersetzt werden.

Whales at risk in sonar sea exercises
Recently, a US judge banned the American Navy from testing a similar system to that which the MoD is keen to introduce. The judge concluded that the booming sounds could damage marine life, yet his comments have done little to deter Britain from entering the low-frequency race in which powerful speakers on a metal post are lowered into the sea. An intense burst of noise designed to detect enemy vessels floods the ocean, causing panic among whales, which use similar sonic booms to find food and mating partners.

(Anzahl Wörter: 95)

A2 Semi-structured interviews

Two interviews were conducted at the usability lab. For the bidirectional translators, the initial interview was conducted after they had performed their first translation task and the final interview after they had commented on their second task. For the unidirectional translators, the initial interview was conducted after they had performed their translation task and the final interview after they had commented on the task.

Initial interview template

Initial Interview

Today's date:

____/____/_____

Assigned code:
SPSS_AC

Prof…….

Initials:

1. Name: _____
SPSS_V1
2. Email address: _____
SPSS_V2
3. Year of birth: _____
SPSS_V3
4. Sex: 1☐ Female 2☐ Male
SPSS_V4 (1-2)

5. Languages acquired/learned?
SPSS_V5_kg_CH-G(), V5_p_CH-G(), V5_hs_CH-G (), V5_c/u_CH-G()
SPSS_V5_kg_G(), V5_p_G(), V5_hs_G(), V5_c/u_G()

Age	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23+
	(kindergarten)						*(primary)*						*(high school)*					*(college/university)*					
1 CH-G																							
2 **German**																							
3 CH-F																							
4 F dialect																							
5 **French**																							
6 CH-I																							
7 I dialect																							
8 **Italian**																							
9 E dialect																							
10 **English**																							
11 Sp dialect																							
12 **Spanish**																							
13																							
14																							
15																							

6. What language do you consider your first language?
SPSS_V6 (1-15)

7. What is your parents' first language?
SPSS_V7 (1-2)
1☐ Mother: _____ 2☐ Father: _____
(1-15) (1-15)

8. Language biography *(how did you learn your languages?)*
SPSS_V8(1-10)
1☐ family 2☐ school 3☐ private language lessons 4☐ with friends/pen friend 5☐ tandem learning 6☐ relatives abroad
7☐ au-pair work 8☐ student exchange 9☐ abroad 10☐ other: _____
10_t (string)

9. What language versions do you work with?
SPSS_V9 (1-6)
1☐ English – German 2☐ German – English 3☐ German – French 4☐ German – Italian
SPSS_V9_1 (1-2)
5☐ other: _____ 6☐ other: _____

10. Education
SPSS_V10 (1-5)
1☐ translation degree: _____ 2☐ other undergraduate degree: _____
1_t (string) 2_t (string)
3☐ translation certificate: _____ 4☐ graduate degree: _____
3_t (string) 4_t (string)
5☐ other: _____
5_t (string)

11. How long have you been working at XY / as a translator?
SPSS_V11_t (1-2)
1☐ full time: _____ 2☐ part time: _____
1_t (string) 2_t (string)

12. What do you normally do at XY / as tasks?
SPSS_V12_t (1-5)
1☐ translation only 2☐ mainly translation, some revision 3☐ half translation, half revision
4☐ mainly revision, some translation 5☐ revision only

13. Other work experience?
SPSS_V13_t (string)

Initial_interview_GE_lab_huna.doc Page: 1/2

14. Usual surrounding:
SPSS_V14 (1-4)
☐ alone (1) ☐ shared office 2-4 (2) ☐ shared office 5-10 (3) ☐ other: _____ (4_t (string))

15. Usual mode:
SPSS_V15 (1-5)
1☐ desktop 2☐ laptop 3☐ handwriting 4☐ handwriting & desktop 5☐ handwriting & laptop

16. Writing by hand: (when and why)
SPSS_V16 (1-2)
1(1-6)☐ Yes 1☐ receive ST on paper 2☐ required 3☐ prefer paper form 4☐ work on train
5☐ only for revision 6☐ other: _____ (6_t (string))
2 (1-6)☐ No 1☐ faster typing 2☐ not used to it 3☐ only work on laptop 4☐ only work on desktop
5☐ waste of time 6☐ other: _____ (6_t (string))

17. Do you consider yourself a touch typist?
SPSS_V17 (1-3) 1☐ definitely 2☐ rather yes 4☐ rather no 3☐ not at all

18. Which operating system do you use?
SPSS_V18 (1-3)
1☐ Mac 2☐ Windows 3☐ other _____ (3_t (string))

19. Do you use CAT tools? Which / why not?
SPSS_V19 (1-2)
1(1-3)☐ Yes 1☐ StarTransit 2☐ SDL Trados 3☐ other: _____ (3_t (string))
2(1-6)☐ No 1☐ don't know any tools 2☐ don't like them 3☐ too complicated 4☐ never had a chance
5☐ don't know how 6☐ other: _____ (6_t (string))

20. What sort of reference tools (Nachschlagewerke) do you use?
SPSS_V20 (1-3)
paper 1☐ never 2☐ rarely 3☐ occasionally 4☐ often 5☐ very frequently 6☐ which: _____
SPSS_V20_1 (1-6) (6_t (string))
electronic 1☐ never 2☐ rarely 3☐ occasionally 4☐ often 5☐ very frequently 6☐ which: _____
SPSS_V20_2 (1-6) (6_t (string))
web-based, online 1☐ never 2☐ rarely 3☐ occasionally 4☐ often 5☐ very frequently 6☐ which: _____
SPSS_V20_3 (1-6) (6_t (string))

21. Describe your typical work patterns: *(start, pause, drafts, proofreading…)*
SPSS_V21_t (string)

22. If you encounter a word or term you don't know – what do you do first? And then?
SPSS_V22_t (string)

23. What kind of text types do you usually translate [G-E]?

24. What kind of text types do you usually translate [E-G]?

25. Handedness 1☐ right-handed 2☐ left-handed
3☐ mouse on right-hand side 4☐ mouse on left-hand side

26. Keyboard 1☐ QWERTZ (German) 2☐ QWERTY (English)

27. Comments:
SPSS_V23_t (string)

Initial_interview_GE_lab_huna.doc Page: 2/2

Final interview template

Final Interview

Name:

Date: ___/___/___
Assigned code: __
Initials:
Version:

Experience with Camtasia and lab situation
Psychological effects (please describe, if any):
Text processing effects (please describe, if any):
Experience with process research
Did you notice anything about your own translation process when watching the recordings?
For bidirectional translators: Do you do anything differently when you're translating into your L1 compared to translating into your L2 or L3?
Comments/other information:

Final_interview_GE_lab_huna.doc 1 / 1

A3 Introduction of participant and workplace set-up

When a participant arrived, she was greeted by a researcher and a technician. For the Bidir group, the researcher (myself) and the technician were always the same. For the comparison groups UniEnglish and UniGerman, whose processes were recorded within the Capturing Translation Processes project, the researcher and the technician varied based on availability. The researcher roughly outlined the procedure and the participant was then taken into the testing room in which the single workplace with a desktop computer, the eye-tracking monitor, a computer mouse and an adjustable chair had been prepared. The same version of a *Windows* computer operating system was kept throughout the whole study so as to ensure comparability of results. The technician followed a script with instructions and made sure the procedure was identical for every participant. She adjusted the office chair to suit the participant and placed the mouse on the side the participant wanted. Then, the technician started the calibration software and the result was checked before she gave the okay to proceed. The technician then went to the control room, which was separated from the testing room by a one-way mirror, and started the screen recording software, the keylogger and the eye-tracking program. On the participant's computer screen, the question of an initial acclimatisation research task appeared: In the English-German task, she had to do internet research to ascertain the number of different tomato species and in the German-English, she had to search for the size of the largest sea.[158] The actual goal of this warm-up task was to help the participant become familiar with the equipment and to overcome any potential initial nervousness. After the participant had presented an answer or about two minutes had passed, the technician started another routine whereby a prompt saying that they could start the translation task by pressing the space bar.[159] When the participant pressed the space bar, the computer desktop appeared with a Microsoft Word document that contained the brief and the ST. Throughout the recording, the technician and the researcher remained in the control room. Whenever the technician was addressed by

158 The original instruction for the Englisch-German task was: "Kurzrecherche: Wieviele Tomatensorten gibt es ungefähr? Mit der Leertaste gelangen Sie zum Internetbrowser." For the G-E task, the instruction read: "Kurzrecherche: Bitte nennen Sie die Oberfläche des grössten Meeres der Welt. Mit der Leertaste gelangen Sie zum Internetbrowser."

159 For the unidirectional translators, whose processes had already been recorded in the CTP project, the instruction read: "Beginnen Sie nun mit der Übersetzung und schreiben Sie den Zieltext bitte in das Dokument1. Mit der Leertaste gelangen Sie zum Ausgangstext". When they pressed the space bar, the computer screen appeared as they had left it after the warm-up task.

a participant with regard to a problem, she gave assistance. She was also the one who started and stopped all the data collection software.

Group	Translation version	Translation direction	Participants	Group size	Translation task	Source text
German bidirectional translators	German-English	into L2	BiDir 1-6	n=6	BiDir_L2	Wale
	English-German	into L1			BiDir_L1	Whales
English unidirectional translators	German-English	into L1	UniEnglish 1-6	n=6	Uni-English_L1	Wale
German unidirectional translators	English-German	into L1	UniGerman 1-6	n=6	Uni-German_L1	Whales

Quasi-experimental design of main study: Participant groups and translation tasks

A4 Using eye tracking as a data collection method

General limitations of eye tracking

One disadvantage is that the whole process has to be performed at a computer – whereas some participants state that they for example revise their target texts on paper – and touch-typing is recommended as eye-tracking data can only be collected when the participant is actually looking at the monitor.

Although the eye tracker is calibrated before starting the experiment, it sometimes happens that it becomes inaccurate during the translation task. Possible causes are that the participants change either their position in front of the screen or, if they wear glasses, the angle at which they gaze at the screen. On the screen and subsequently during retrospection this can be noticed as a shift of the gaze pattern between two lines of displayed text. As the intention of this study was never to analyse gaze patterns on the word level, this shift did not cause distortions of the data. The comments during retrospection show that the participants also understood the visualisation of their eye movements as general cues to what they had been doing while translating.

If eye-tracking measures had been an integral part of this study, much of its claim to ecological validity would have had to be sacrificed since participants would not have been allowed to move their heads or to freely arrange the windows on their screen. All windows would have needed to be fixed (e. g. ST on the top left, TT on the top right, internet browser on the bottom) in order to establish stable areas of interest for the eye-tracking measures.

Settings of the eye tracker

In the present study, the fixation filter for the eye tracker was set to 50 pixels as the fixation radius and to 100ms as the minimum fixation duration, which is a threshold often used in translation process research. This means that when at least six succeeding gaze samples are recorded within 50 pixels from each other, they are considered part of the same fixation. This setting is comparable to the one tested and proposed by the EYE-to-IT project (Jakobsen and Jensen 2008) [160] and by Hvelplund (2011), who used 40 ms and 100 pixels with a 50-Hertz eye tracker.

160 If not indicated otherwise, see Chapter 10 Bibliography for complete reference.

Appendix B: Substudy on product quality

This appendix contains information on the substudy (see Section 4.3 of the chapter *Methodology*), which investigated the translation products. Its main component was an evaluation task performed by 36 readers of the English target texts. Therefore, the participant data is repeated here and the template for the evaluation task is made available. The stimuli for the task, i. e. the English texts are reproduced in Appendix D.

Participant data

Group	Translation version	Translation direction	Participants	Group size	Translation task	Source text
German bidirectional translators	German-English	into L2	BiDir 1-6	n=6	BiDir_L2	Wale
	English-German	into L1			BiDir_L1	Whales
English unidirectional translators	German-English	into L1	UniEnglish 1-6	n=6	UniEnglish_L1	Wale
German unidirectional translators	English-German	into L1	UniGerman 1-6	n=6	UniGerman_L1	Whales

Data collected in substudy

Part of substudy	Type of information	Data collection instrument	Form of data	Type of data
Product descriptors	Word and character counts	word count in Microsoft Word	numbers	quantitative
Product quality	Readability of English target texts	Flesch Reading Ease formula	scores	quantitative
	Acceptability of English target texts	calculation based on results from ranking task	weighted scores	quantitative
	Personal background of raters	questionnaire	transcript	qualitative
	Raters' individual evaluation criteria	questionnaire	transcript	qualitative

Template for the ranking task

36 representatives of the intended target audience, i. e. news(paper) readers, evaluated the twelve English target texts according to their own criteria. They then had to rank the texts and fill in a questionnaire.

Ranking twelve journalistic texts

Dear participant
The following twelve texts were produced by professional translators. They had been asked to translate a short text for a quality English-language newspaper. You, as a potential reader of that newspaper, can now decide which of the twelve texts are the three most acceptable ones and which are the three least acceptable ones.
Please do not change the order of the texts. Please work on your own and do not look anything up but feel free to annotate the texts. After having decided on the ranking, please fill in the ranking sheet and the background information on the last two pages. All information will be treated anonymously and is only used for research purposes.
This task will take you about half an hour.

Thank you very much for your support.
Kind regards

Andrea Hunziker Heeb
Institute of Translation and Interpreting
Zurich University of Applied Sciences
andrea.hunziker@zhaw.ch

My supervisor Prof. Maureen Ehrensberger-Dow is also available for any questions you might have on the project at ehre@zhaw.ch.

Document code: 234VD

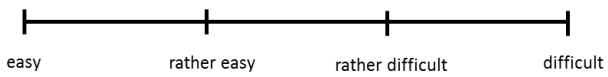
[On the next twelve pages, the twelve English target texts were printed, each fitted with a document code so that the raters could identify neither the translator nor the translation direction]

Ranking

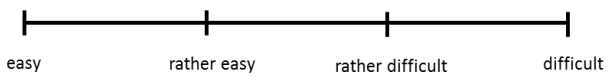
Please enter the codes of the texts you chose. You find the code on the same page as the respective text.

Rank	Document code	
1.		most acceptable
2.		
3.		
⋮	⋮	
10.		least acceptable
11.		
12.		

1. How easy or difficult was it to rank the three most acceptable texts? Please indicate by placing a mark on the line below.



2. How easy or difficult was it to rank the three least acceptable texts? Please indicate by placing a mark on the line below.



3. What criteria did you use for your ranking? (Feel free to write in English, French, German, Italian or Spanish)

ZHAW_Hunziker Heeb_Set 1 Page: 14/15

4. If you have any comments concerning the task, please enter them here (Feel free to write in English, French, German, Italian or Spanish):

Background information

5. What language(s) did you speak before entering school?

_____

_____

6. At what age did you start to learn English at school? _____

7. What languages do you speak? Please list them by starting with your strongest one.

1) _____ 4) _____

2) _____ 5) _____

3) _____ 6) _____

8. Towards which degree do you study?

_____

9. Which year are you in?

_____

10. Have you ever had a course in text revision or editing?

□ yes □ no

11. How frequently do you read English-language newspapers (online/printed)?

□ daily □ about once a week □ about once a month □ less than once a month

12. Year of birth: _____

13. Sex: □ female □ male □ not specified

Thank you!

ZHAW_Hunziker Heeb_Set 1 Page: 15/15

Appendix C: Results for translators' processes

This appendix contains information on the preparation and analysis of data to investigate the translation processes, which was done in the main study (see Section 4.2 of the chapter *Methodology*). It also contains all the results that are presented in Chapter 5 *The translators' processes*. For the reader's convenience, I decided to include the tables and graphs of all the comparisons, i. e. those that yielded significant differences as well as those that did not. The comparisons are between the task groups of the bidirectional translators working into their L2 (Bidir_L2) and into their L1 (Bidir_L1), the unidirectional English translators (UniEnglish_L1) and the unidirectional German translators (UniGerman_L1) working into their respective L1s.

C1 Data preparation and data analysis of main study

All translation processes were transcribed based on the definitions for the four process activities used in this study (see Section 4.2.6 in Chapter *Methodology* for operationalisations). Information on how I proceeded is given below.

Use of keystroke logs

The version of the keystroke-logging software Inputlog that I used to log the bidirectional translators' processes provides a general report, which served as my main source for process transcription (for an excerpt see below). Additionally, Inputlog offers some default automatic reports as e. g. for the number, content and duration of text revisions. However, it was not possible to use any of them one to one because of differing definitions of process activities. For example, translators tend to paste the source text into the target text file and then produce the target text above that source text or replace single source text sentences with target text. I transcribed those activities as writing or revision activities whereas Inputlog categorised them as revisions only. Inputlog had been developed to analyse writing processes and assumes that writers start from a blank page.

The start and end times of the logged key and mouse activities in the output file corresponded to the time passed since the program had been started and thus needed to be adjusted to represent translation process times. As a second data source, I used the screen videos with integrated gaze data visualisation was used, e. g. to determine the focus of attention during pauses or to trace external interruptions. To extract information of the two data sources, time stamps needed to be synchronised. One one hand, these

necessary synchronisations were time-consuming, on the other hand, the use and triangulation of multiple data sources facilitated experimental rigor and led to new insights as for example behaviour during long pauses (see indicator of coordination effort in Section 7.4).

id	type	output	position-Full	doc-lengthFull	char-Production	startTime	startClock	endTime	endClock	actionTime	pauseTime	pause-Location	pauseLocationFull	x	y	position	doc-length
408	mouse	Movement	0	0	0	322017	00:05:22.017	323483	00:05:23.483	1466	1092	13	UNKNOWN	522	268		
409	keyboard	LSHIFT	0	747	0	327617	00:05:27.617	327960	00:05:27.960	343	4134	2	BEFORE WORDS			0	747
410	keyboard	S	0	747	747	327867	00:05:27.867	327929	00:05:27.929	62	250	2	BEFORE WORDS			0	747
411	replacement	[0211] S	1	727	747	327867	00:05:27.867	327939	00:05:27.939	0	0	1	WITHIN WORDS			1	727
413	keyboard	t	1	727	747	328085	00:05:28.085	328132	00:05:28.132	47	0	1	WITHIN WORDS			1	727
414	keyboard	r	2	728	747	328226	00:05:28.226	328288	00:05:28.288	62	141	1	WITHIN WORDS			2	728
415	keyboard	a	3	729	748	328335	00:05:28.335	328382	00:05:28.382	47	109	1	WITHIN WORDS			3	729
416	keyboard	n	4	730	749	328491	00:05:28.491	328553	00:05:28.553	62	156	1	WITHIN WORDS			4	730
417	keyboard	d	5	731	750	328600	00:05:28.600	328647	00:05:28.647	47	109	1	WITHIN WORDS			5	731
418	keyboard	i	6	732	751	328818	00:05:28.818	328881	00:05:28.881	63	218	1	WITHIN WORDS			6	732
419	keyboard	n	7	733	752	329037	00:05:29.037	329084	00:05:29.084	47	219	1	WITHIN WORDS			7	733
420	keyboard	g	8	734	753	329130	00:05:29.130	329177	00:05:29.177	47	93	1	WITHIN WORDS			8	734
421	keyboard	SPACE	9	735	754	329255	00:05:29.255	329302	00:05:29.302	47	125	3	AFTER WORDS			9	735
422	keyboard	o	10	736	755	329411	00:05:29.411	329458	00:05:29.458	47	156	2	BEFORE WORDS			10	736
423	keyboard	f	11	737	756	329567	00:05:29.567	329630	00:05:29.630	63	156	1	WITHIN WORDS			11	737
424	keyboard	SPACE	12	738	757	329723	00:05:29.723	329786	00:05:29.786	63	156	3	AFTER WORDS			12	738
425	keyboard	w	13	739	758	329879	00:05:29.879	329942	00:05:29.942	63	156	2	BEFORE WORDS			13	739
426	keyboard	h	14	740	759	330004	00:05:30.004	330066	00:05:30.066	62	125	1	WITHIN WORDS			14	740
427	keyboard	a	15	741	760	330082	00:05:30.082	330144	00:05:30.144	62	78	1	WITHIN WORDS			15	741
428	keyboard	l	16	742	761	330222	00:05:30.222	330269	00:05:30.269	47	140	1	WITHIN WORDS			16	742
429	keyboard	e	17	743	762	330363	00:05:30.363	330410	00:05:30.410	47	141	1	WITHIN WORDS			17	743
430	keyboard	s	18	744	763	330488	00:05:30.488	330550	00:05:30.550	62	125	1	WITHIN WORDS			18	744
431	mouse	Movement	18	744	764	332079	00:05:32.079	333764	00:05:33.764	1685	1591	1	WITHIN WORDS	331	294		
432	mouse	Movement	18	744	764	335994	00:05:35.994	336135	00:05:36.135	141	2230	1	WITHIN WORDS	331	296		
433	keyboard	LALT	19	745	764	338179	00:05:38.179	338335	00:05:38.335	156	2044	2	BEFORE WORDS			19	745
434	keyboard	TAB	19	745	765	338257	00:05:38.257	338444	00:05:38.444	187	78	2	BEFORE WORDS			19	745
435	focus	Programmumschaltung	19	745	765	338288	00:05:38.288	338288	00:05:38.288	0	0	10	TRANSITION				
436	mouse	Movement	19	745	765	338288	00:05:38.288	338288	00:05:38.288	0	0	13	UNKNOWN	331	297		
437	focus	grounding of whales - Google-Suche - Mozilla Firefox	19	745	765	338444	00:05:38.444	338444	00:05:38.444	0	0	10	TRANSITION				
438	mouse	Movement	19	745	765	338834	00:05:38.834	339723	00:05:39.723	889	0	13	UNKNOWN	388	19		
439	mouse	LEFT Click	19	745	765	339832	00:05:39.832	339926	00:05:39.926	94	109	2	BEFORE WORDS	388	19		
440	focus	stranding - Englisch - Deutsch Wörterbuch - leo.org - Mozilla Firefox	19	745	765	339926	00:05:39.926	339926	00:05:39.926	0	0	10	TRANSITION				
441	keyboard	h	19	745	765	340534	00:05:40.534	340581	00:05:40.581	47	0	1	WITHIN WORDS				
442	keyboard	a	19	745	765	340628	00:05:40.628	340690	00:05:40.690	62	94	1	WITHIN WORDS				
443	keyboard	n	19	745	765	340721	00:05:40.721	340784	00:05:40.784	63	93	1	WITHIN WORDS				
444	keyboard	g	19	745	765	340877	00:05:40.877	340940	00:05:40.940	63	156	1	WITHIN WORDS				
445	keyboard	SPACE	19	745	765	341049	00:05:41.049	341111	00:05:41.111	62	172	3	AFTER WORDS				
446	keyboard	z	19	745	765	341205	00:05:41.205	341252	00:05:41.252	47	156	2	BEFORE WORDS				
447	keyboard	u	19	745	765	341330	00:05:41.330	341377	00:05:41.377	47	125	5	AFTER SENTENCES				
448	keyboard	RETURN	19	745	765	341595	00:05:41.595	341642	00:05:41.642	47	265	4	BEFORE SENTENCES				

Excerpt from keystroke log file for Bidir2_L2

Transcription

For the Bidir processes, I transcribed the KSL output files using the four main categories *writing*, *revising*, *information seeking* and *pausing*. Their operationalisation is described in the chapter *Methodology*. An excerpt of such a transcript can be seen in the following table. The two rows highlighted in yellow are the transcription of the whole sequence of activities from the previous table (marked with id 408 to id 448).

activity	activity subtype	start time (ms)	end time (ms)	duration (ms)	pause location	content of writes/revises	characters and spaces	words	search topic
writing	first	327617	338288	10671		Stranding·of·whales	19	3	
info seek.		338444	355947	17503					hang zu
pausing		362234	367553	5319	after search				
writing		367553	371251	3698		The·phenomenon·is·	18	3	
pausing		371188	377163	5975	between writing acts.				
writing		377163	386180	9017		unlikely·to·be·linked·to·a·	27	6	
info seek.		386180	392982	6802					hang zu
revising	deletes	398364	398442	78		·	1	0	
writing		399971	400220	249		n·	2	0	
info seek.		403949	427364	23415					hang zu

Transcript of excerpt from keystroke log for Bidir2_L2

The unidirectional translators' TPs recorded and commented on in the usability laboratory had already been transcribed using XML-mark-up according to conventions developed and refined in the CTP project (cf. Massey and Ehrensberger-Dow 2014, 85), which are based on the guidelines by the Text Encoding Initiative[161]. Location and lengths of screen activities had been measured by the clock on the screen videos. Since I applied different definitions of pauses and online information searches, I extracted the information I needed with the Transvis process visualiser[162] and merged it with information from the keystroke logs. I spot-checked all 24 transcripts for my study to ensure experimental rigor and then analysed them using Microsoft Excel.

161 The Text Encoding Initiative (TEI) Guidelines for Electronic Text Encoding and Interchange specifies methods for marking up machine-readable texts. More information is available at http://www.tei-c.org.

162 Transvis is a semi-automated program to visualise translation processes from XML transcripts based on screen recordings and generate reports. The program and its open-source code are available at https://github.com/sykaeh/TransVis.

The retrospective comments by all participants, were also transcribed following the TEI guidelines. Those transcripts are in XML format.

Interruptions in translation processes

Interruptions occurred in nine of the 24 translation processes. For example, on one occasion, the technician needed to restart the logging software, and in another, she had to provide login information for the translator to be able to access online resources. If an external interruption had occurred during the translation processes, the duration and number of keystrokes involved were deducted from the total numbers. Although the durations could be determined for the actual interruptions, the time a translator might have needed for reorientation after the incident could not be established and therefore is included in the total time on task.

The following table lists all traceable interruptions during the recording of the translation processes. In order to calculate their duration as accurately as possible, the screen recordings including the eye-gaze data, the keystroke logs and the retrospective verbalisations were consulted.

Process	Start interruption in keystroke log (ms)	End interruption in keystroke log (ms)	Duration (sec)	Source of interruption	Location in TP
Bidir1_L2	617532	618813	1,3	window pops up with suggestion to save file	drafting phase
Bidir1_L2	719282	738391	19,1	technician has to do login to access dictionary	drafting phase
Bidir5_L2	624359	626203	1,8	window pops up with suggestion to save file	drafting phase
Bidir5_L2	676047	703562	27,5	technician has to do login to access dictionary	drafting phase
Bidir6_L2	619500	623625	4,1	window pops up with suggestion to save file	drafting phase
Bidir1_L1	107531	112453	4,9	black instruction screen pops up late	orientation phase
Bidir1_L1	635203	642906	7,7	window pops up with suggestion to save file	drafting phase
Bidir4_L1	294327	321611	27,3	technician explains how to type „Ä“	drafting phase
Bidir6_L1	619109	625500	6,4	window pops up with suggestion to save file	drafting phase
UniEnglish1_L1	134266	147500	13,2	technician checks whether programs are still running after tr used a certain keyboard shortcut	orientation phase
UniGerman2_L1	236797	267328	30,5	technician shows translator how to open an additional browser tab	drafting phase
UniGerman4_L1	356922	405281	48,4	technician has to make cursor reappear on screen	drafting phase

List of interruptions during translation tasks

C2 Global process measures

All translation processes were analysed with regard to their duration and the number of characters produced.

Process duration

Process	Process duration (sec)	Process	Process duration (sec)
Bidir1_L2	1586	Bidir1_L1	1136
Bidir2_L2	1708	Bidir2_L1	1649
Bidir3_L2	2842	Bidir3_L1	1545
Bidir4_L2	1335	Bidir4_L1	1586
Bidir5_L2	2283	Bidir5_L1	2049
Bidir6_L2	2510	Bidir6_L1	2085

Duration of Bidir_L2 and Bidir_L1 translation processes (in seconds) ordered by translator

Process	Process duration (sec)	Process	Process duration (sec)
UniEnglish1_L1	640	Bidir4_L2	1335
UniEnglish5_L1	922	Bidir1_L2	1586
UniEnglish6_L1	1081	Bidir2_L2	1708
UniEnglish2_L1	1318	Bidir5_L2	2283
UniEnglish3_L1	1396	Bidir6_L2	2510
UniEnglish4_L1	1407	Bidir3_L2	2842

Duration of UniEnglish_L1 and Bidir_L2 processes in ascending order (in seconds)

Process	Process duration (sec)	Process	Process duration (sec)
UniGerman4_L1	887	Bidir1_L1	1136
UniGerman2_L1	1021	Bidir3_L1	1545
UniGerman6_L1	1126	Bidir4_L1	1586
UniGerman3_L1	1157	Bidir2_L1	1649
UniGerman5_L1	1165	Bidir5_L1	2049
UniGerman1_L1	1896	Bidir6_L1	2085

Duration of UniGerman_L1 and Bidir_L1 processes in ascending order (in seconds)

Character count

Process	Character count	Process	Character count
Bidir1_L2	1185	Bidir1_L1	990
Bidir2_L2	1159	Bidir2_L1	1073
Bidir3_L2	1507	Bidir3_L1	1224
Bidir4_L2	1103	Bidir4_L1	1191
Bidir5_L2	1401	Bidir5_L1	1333
Bidir6_L2	1851	Bidir6_L1	1426

Character count for Bidir_L2 and Bidir_L1 processes (no.) ordered by translator

Process	Character count	Process	Character count
UniEnglish1_L1	829	Bidir4_L2	1103
UniEnglish4_L1	954	Bidir2_L2	1159
UniEnglish2_L1	1074	Bidir1_L2	1185
UniEnglish6_L1	1125	Bidir5_L2	1401
UniEnglish3_L1	1185	Bidir3_L2	1507
UniEnglish5_L1	1381	Bidir6_L2	1851

Character count for UniEnglish_L1 and Bidir_L2 processes (no.) in ascending order

Process	Character count	Process	Character count
UniGerman4_L1	859	Bidir1_L1	990
UniGerman6_L1	949	Bidir2_L1	1073
UniGerman3_L1	1014	Bidir4_L1	1191
UniGerman2_L1	1182	Bidir3_L1	1224
UniGerman5_L1	1204	Bidir5_L1	1333
UniGerman1_L1	1397	Bidir6_L1	1426

Character count for UniGerman_L1 and Bidir_L1 processes in ascending order

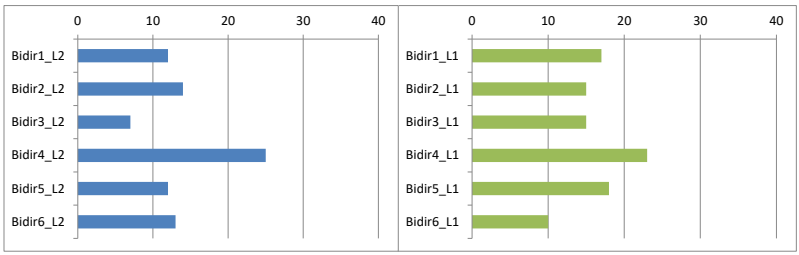
C3 Total numbers of process activities

Process	writing	revising	information seeking	pausing	total number of activities
Bidir1_L2	33	29	16	26	104
Bidir2_L2	41	72	21	30	164
Bidir3_L2	35	75	25	48	183
Bidir4_L2	55	72	14	20	161
Bidir5_L2	47	74	29	32	182
Bidir6_L2	55	76	39	32	202
median	44	73	23	31	173
Bidir1_L1	33	37	11	26	107
Bidir2_L1	41	48	17	31	137
Bidir3_L1	38	39	11	36	124
Bidir4_L1	60	112	20	20	212
Bidir5_L1	60	108	19	40	227
Bidir6_L1	36	63	32	25	156
median	40	56	18	29	150
UniEnglish1_L1	31	36	6	16	89
UniEnglish2_L1	33	50	4	44	131
UniEnglish3_L1	26	89	10	35	160
UniEnglish4_L1	62	89	18	28	197
UniEnglish5_L1	50	53	13	19	135
UniEnglish6_L1	34	31	19	22	106
median	34	52	12	25	133
UniGerman1_L1	48	78	26	37	189
UniGerman2_L1	43	59	8	23	133
UniGerman3_L1	33	30	8	37	108
UniGerman4_L1	31	31	6	20	88
UniGerman5_L1	41	128	8	24	201
UniGerman6_L1	54	37	16	40	147
median	42	48	8	31	140

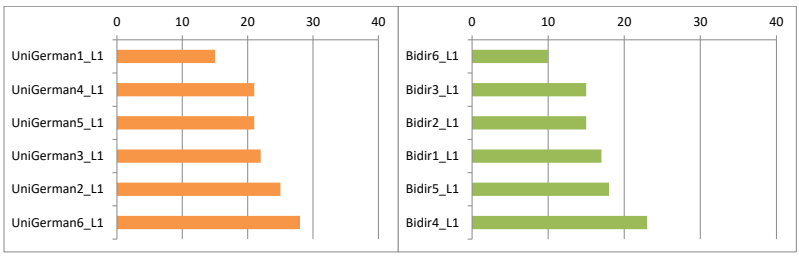
Number of process activities in each translation process and medians per task group

C4 Writing in the translation process

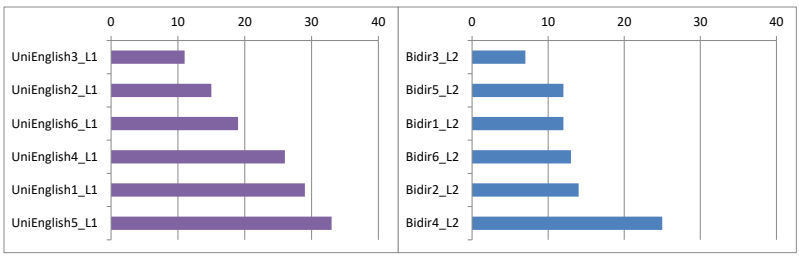
Frequency of writing



Frequency of writing in Bidir_L2 and Bidir_L1 processes ordered by translator (no./10 min)

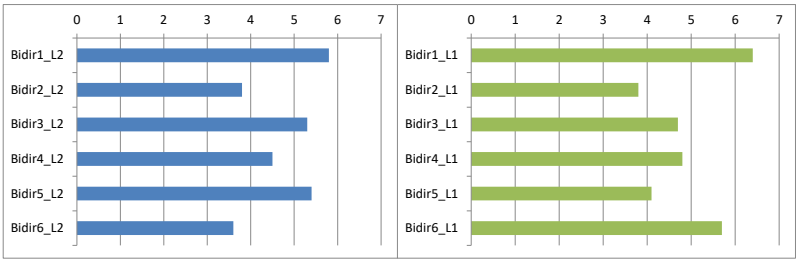


Frequency of writing in UniEnglish_L1 and Bidir_L2 processes in ascending order (no/10 min)

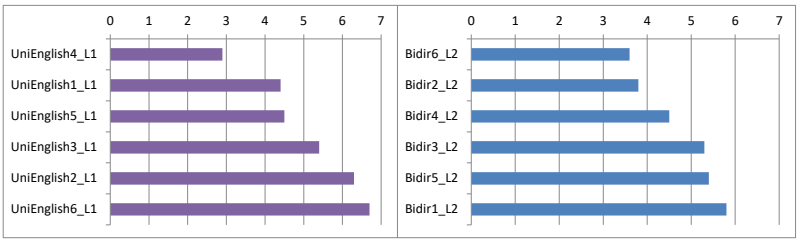


Frequency of writing in UniGerman_L1 and Bidir_L1 processes in ascending order (no/10 min)

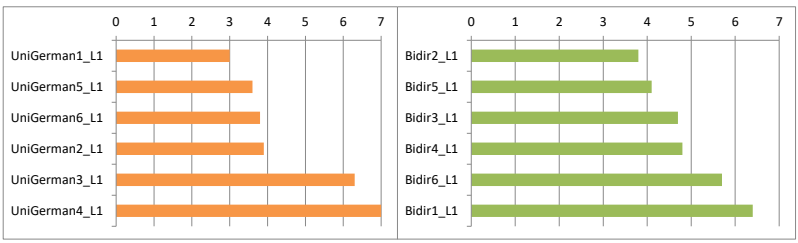
Mean duration and mean length of writing incidents



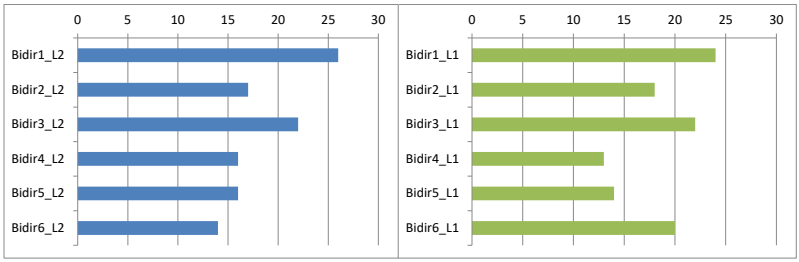
Mean duration of writing incidents (sec) for Bidir_L2 and Bidir_L1 processes, ordered by translator



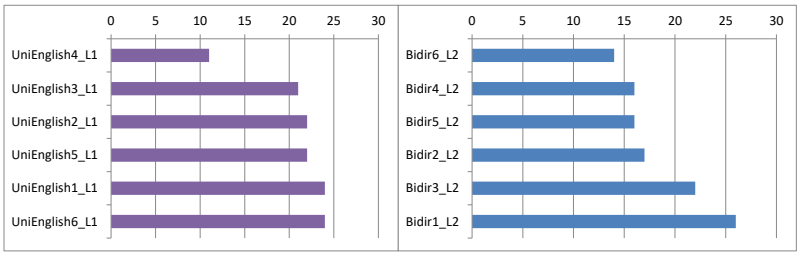
Mean duration of writing incidents (sec) for UniEnglish_L1 and Bidir_L2 processes, in ascending order



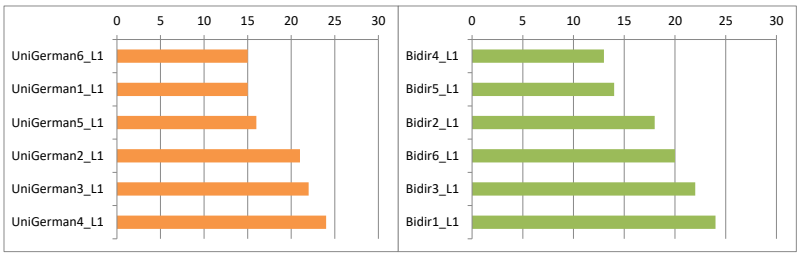
Mean duration of writing incidents (sec) for UniGerman_L1 and Bidir_L1 processes, in ascending order



Mean length of writing incidents (char) for Bidir_L2 and Bidir_L1 processes, ordered by translator



Mean length of writing incidents (char) for UniEnglish_L1 and Bidir_L2 processes, in ascending order



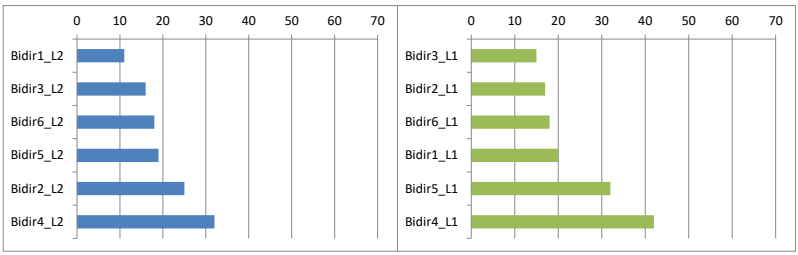
Mean length of writing incidents (char) for UniGerman_L1 and Bidir_L1 processes, in ascending order

Group	Source text	Mean duration of writing sequences (sec)		Mean length of writing sequences (char)	
		range	median	range	median
Bidir_L2	Wale	3.6-5.8	4,9	14-26	16
UniEnglish_L1	Wale	2.9-6.7	4,9	11-24	22
Bidir_L1	whales	3.8-6.4	4,8	13-24	19
UniGerman_L1	whales	3.0-7.0	3,8	15-24	19

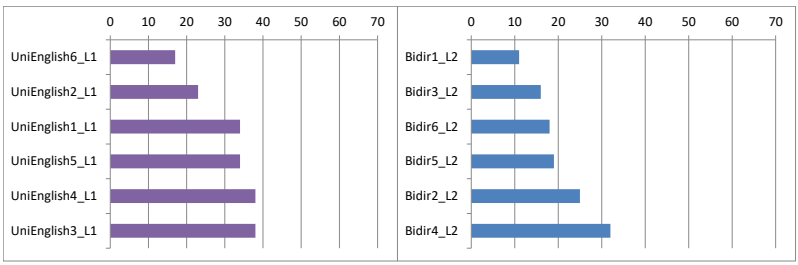
Mean duration of writing incidents (in seconds) and mean length of writing incidents (in characters) per task group

C5 Revising in the translation process

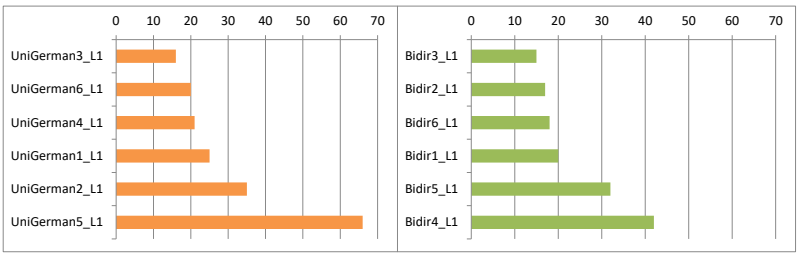
Frequency of revising



Frequency of revising in Bidir_L2 and Bidir_L1 processes ordered by translator (no/10 min)



Frequency of revising in UniEnglish_L1 and Bidir_L2 processes in ascending order (no/10 min)

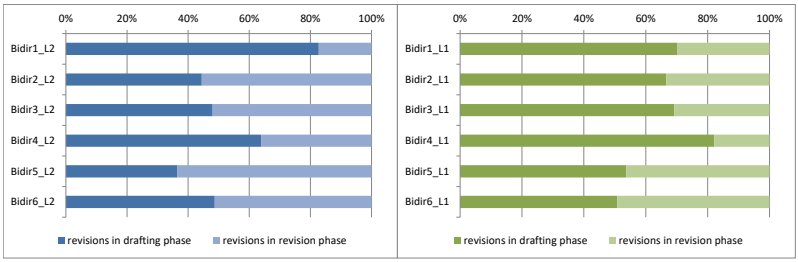


Frequency of revising in UniGerman_L1 and Bidir_L1 processes in ascending order (no/10 min)

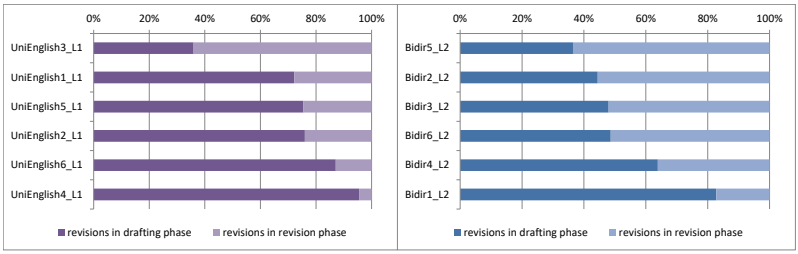
Proportion of revisions per translation process phase

Process	Revisions in drafting phase (%)	Revisions in revision phase (%)
Bidir1_L2	83	17
Bidir2_L2	44	56
Bidir3_L2	48	52
Bidir4_L2	64	36
Bidir5_L2	36	64
Bidir6_L2	49	51
median	48	52
Bidir1_L1	70	30
Bidir2_L1	67	33
Bidir3_L1	69	31
Bidir4_L1	82	18
Bidir5_L1	54	46
Bidir6_L1	51	49
median	68	32
UniEnglish1_L1	72	28
UniEnglish2_L1	76	24
UniEnglish3_L1	36	64
UniEnglish4_L1	96	4
UniEnglish5_L1	75	25
UniEnglish6_L1	87	13
median	76	24
UniGerman1_L1	79	21
UniGerman2_L1	59	41
UniGerman3_L1	60	40
UniGerman4_L1	87	13
UniGerman5_L1	34	66
UniGerman6_L1	70	30
median	65	35

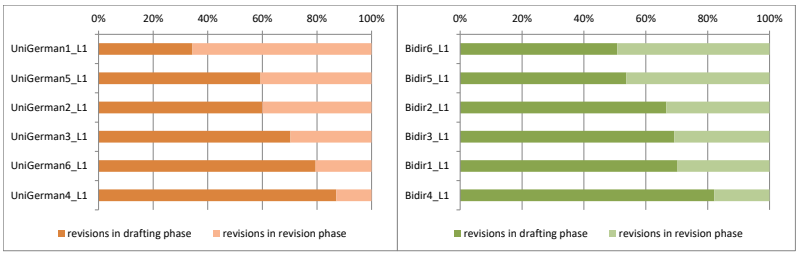
Proportion of revisions performed in translation process phases (%)



Proportion of revisions per TP phase (%) in Bidir_L2 and Bidir_L1 processes, ordered by translator.

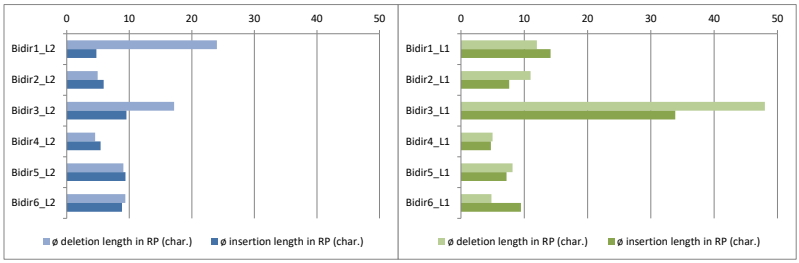


Proportion of revisions per TP phase (%) in UniEnglish_L1 and Bidir_L2 processes, in ascending order by revisions in drafting phase.

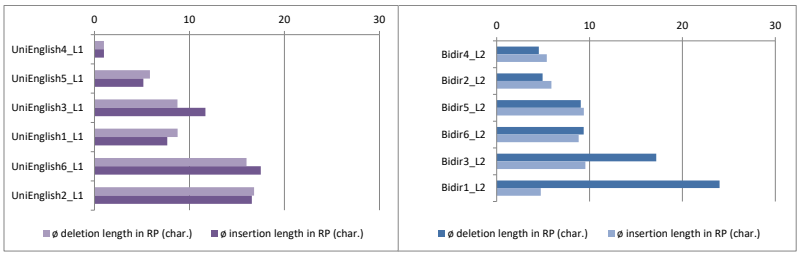


Proportion of revisions per TP phase (%) in UniGerman_L1 and Bidir_L1 processes, in ascending order by revisions in drafting phase.

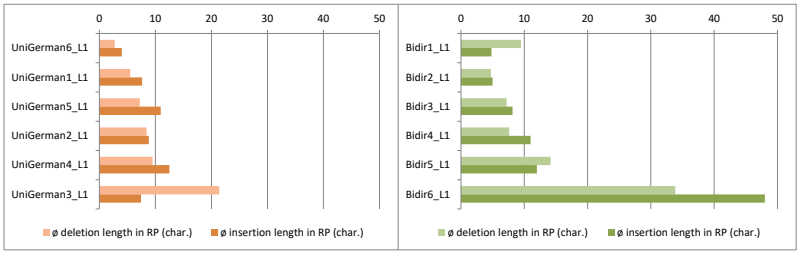
Mean length of revisions in revision phase



Mean length of deletions and insertions (in characters) in the revision phase of Bidir_L2 and Bidir_L1 processes, ordered by translator



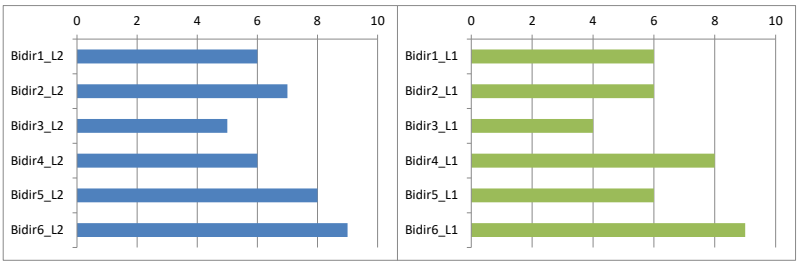
Mean length of deletions and insertions in the revision phase of UniEnglish_L1 and Bidir_L2 processes (in characters), in ascending order by the category *deletion length*



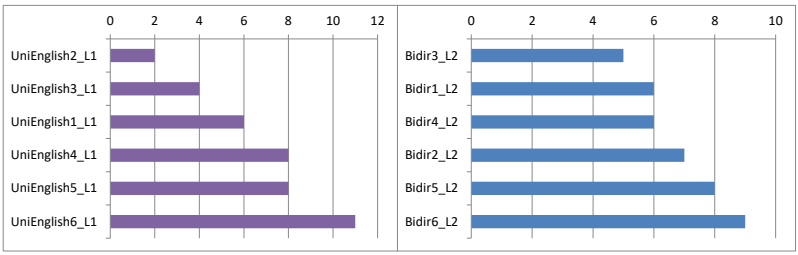
Mean length of deletions and insertions in the revision phase of UniGerman_L1 and Bidir_L1 processes (in characters), in ascending order by the category *deletion length*

C6 Information seeking in the translation process

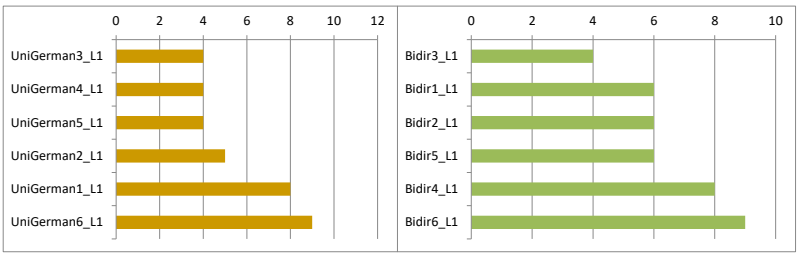
Frequency of information seeking



Frequency of information seeking in Bidir_L2 and Bidir_L1 processes (no/10 min), ordered by translator

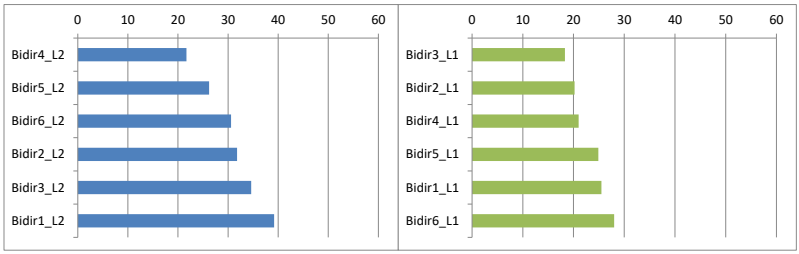


Frequency of information seeking in UniEnglish_L1 and Bidir_L2 processes (no/10 min), in ascending order

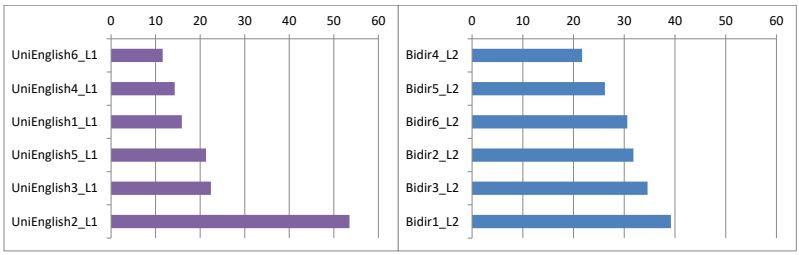


Frequency of information seeking in UniGerman_L1 and Bidir_L1 processes (no/10 min), in ascending order

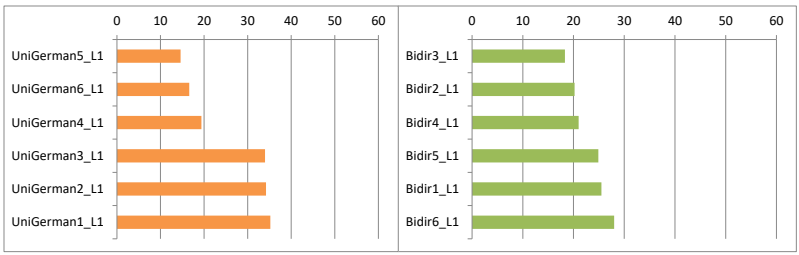
Mean duration of information searches



Mean duration of information searches (sec) for Bidir_L2 and Bidir_L1 processes, ordered by translator

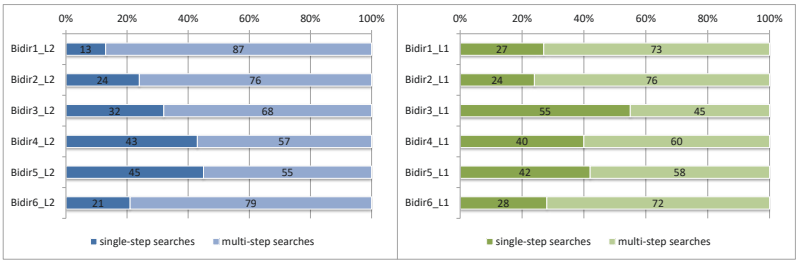


Mean duration of information searches (sec) for UniEnglish_L1 and Bidir_L2 processes, in ascending order

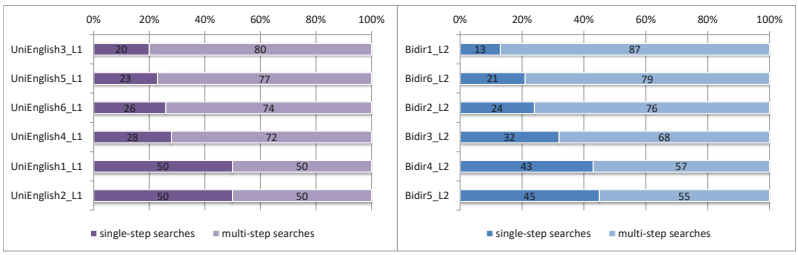


Mean duration of information searches (sec) for UniGerman_L1 and Bidir_L1

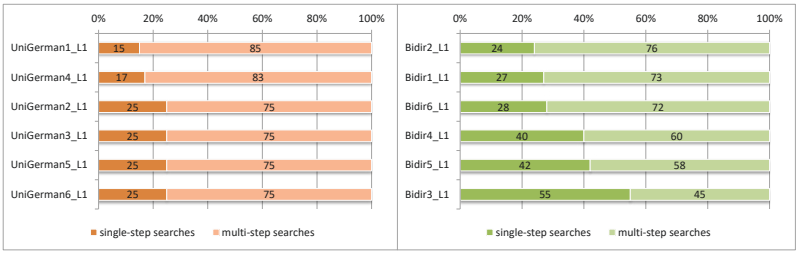
Ratio of single-step to multi-step information searches



Ratio of single-step to multi-step information searches (%) in Bidir_L2 and Bidir_L1 processes ordered by translator

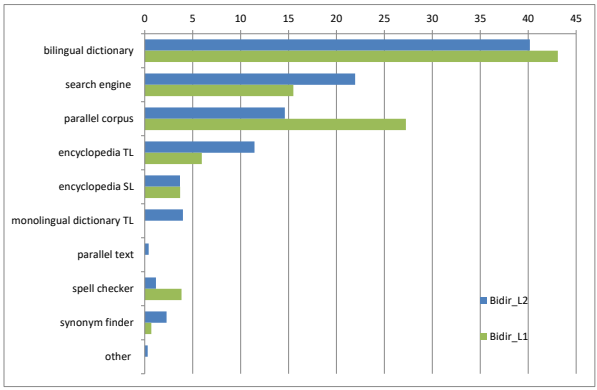


Ratio of single-step to multi-step information searches (%) in UniEnglish_L1 and Bidir_L2 processes, in ascending order by single-step searches

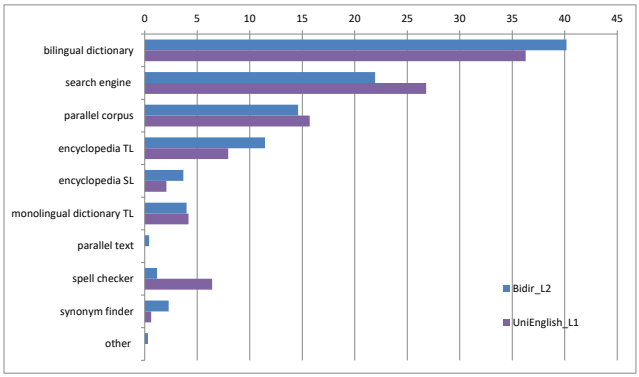


Ratio of single-step to multi-step information searches (%) in UniGerman_L1 and Bidir_L1 processes, in ascending order by single-step searches

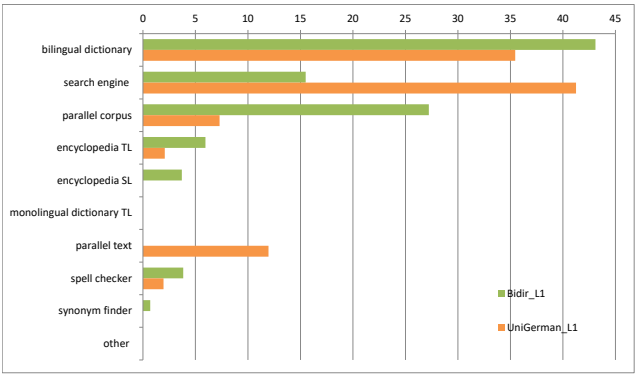
Types of online resources used



Mean use of types of online resources in Bidir_L2 and Bidir_L1 processes (%)



Mean use of types of online resources in Bidir_L2 and UniEnglish_L1 processes (%)



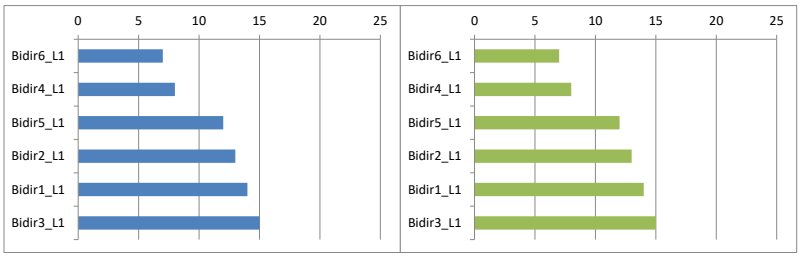
Mean use of types of online resources in Bidir_L1 and UniGerman_L1 processes (%)

Type of resource	Bidir_L2		Bidir_L1		UniEnglish_L1		UniGerman_L1	
	Range	Median	Range	Median	Range	Median	Range	Median
Bilingual dictionary	17.1–67.9	34.7	10.0–75.0	37.0	9.1–85.7	38.9	0.0–85.0	30.2
Search Engine	2.9–42.9	17.2	0.0–63.2	12.2	0.0–44.8	17.0	10.0–60.0	43.6
Parallel corpus	0.0–48.6	5.8	0.0–33.3	32.2	0.0–61.5	13.8	0.0–30.0	25
Encyclopedia TL	0.0–28.6	7.0	0.0–30.0	0.0	0.0–19.7	3.6	0.0–12.5	0.0
Encyclopedia SL	0.0–9.1	3.1	0.0–5.3	0.0	0.0–18.2	1.7	0.0–0.0	0.0
Monolingual dictonary TL	0.0.21.4	0.0	0.0–25.0	0.0	0.0–0.0	0.0	0.0–0.0	0.0
Parallel text	0.0–2.5	0.0	0.0–0.0	0.0	0.0–0.0	0.0	0.0–45.5	0.0
Spell checker	0.0–4.5	0.0	0.0–20.8	4.1	0.0–7.7	3.6	0.0–10.0	0.0
Synonym finder	0.0–13.6	0.0	0.0–3.7	0.0	0.0–4.2	0.0	0.0–0.0	0.0
Other	0.0–1.8	0.0	0.0–0.0	0.0	0.0–0.0	0.0	0.0–0.0	0.0

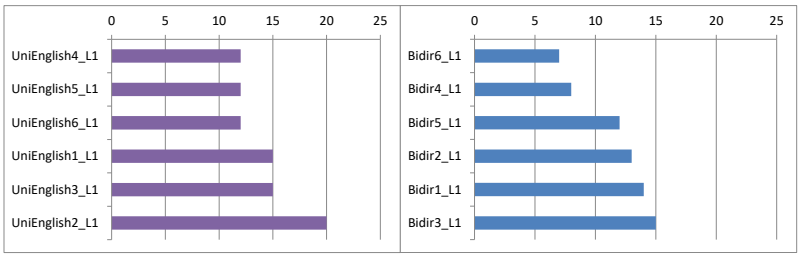
Types of used online resources (%) per task group

C7 Pausing in the translation process

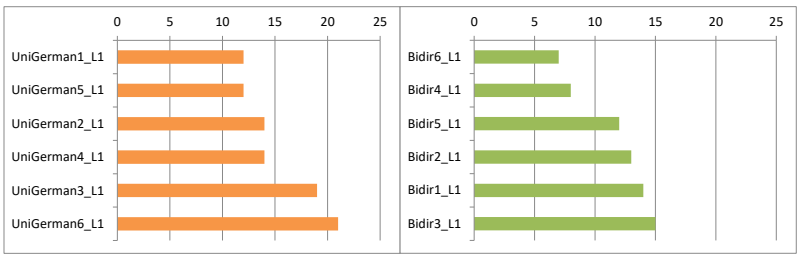
Frequency of pausing



Frequency of pausing in Bidir_L2 and Bidir_L1 processes ordered by translator (no./10 min)

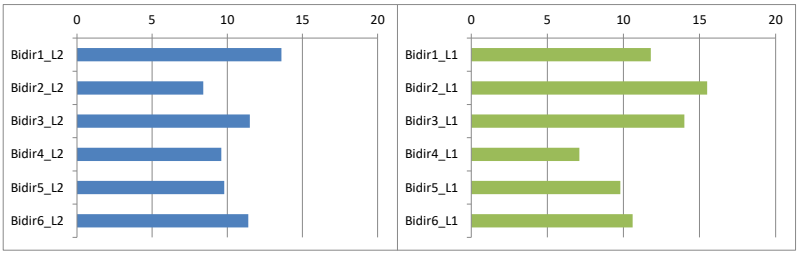


Frequency of pausing in UniEnglish_L1 and Bidir_L2 processes in ascending order (no./10 min)

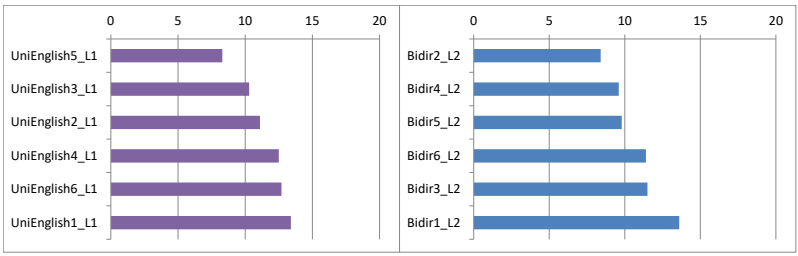


Frequency of pausing in UniGerman_L1 and Bidir_L1 processes in ascending order (no/10 min)

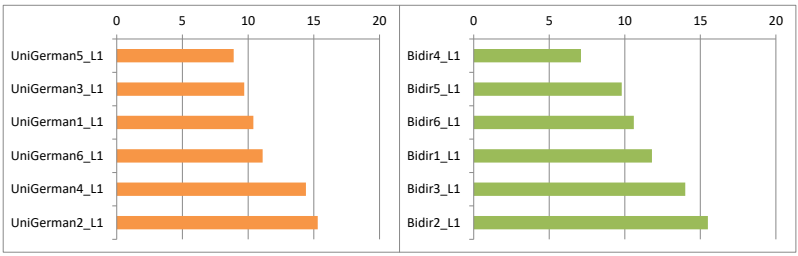
Mean duration of pauses



Mean duration of pauses (sec) for Bidir_L2 and Bidir_L1 processes, ordered by translator



Mean duration of pauses (sec) for UniEnglish_L1 and Bidir_L2 processes, in ascending order



Mean duration of pauses (sec) for UniGerman_L1 and Bidir_L1 processes, in ascending order

Group	Source text	Range of mean duration (sec)	Median of mean duration (sec)
Bidir_L2	Wale	8.4-13.6	10,6
UniEnglish_L1	Wale	8.3-13.4	11,8
Bidir_L1	whales	7.1-15.5	11,2
UniGerman_L1	whales	8.9-15.3	10,7

Range and median of mean pause duration (sec) per task group

Overview of all process descriptors and the results of their comparisons between the task groups

No.	Process descriptor	Bidir_L2 vs. Bidir_L1	Bidir_L2 vs. UniEnglish_ L1 (German-English)	Bidir_L1 vs. UniGerman_ L1 (English-German)
1	process duration	n. s.	p<0.05	p<0.05
2	character count	n. s.	n. s.	n. s.
3	total no. of process activities	n. s.	n. s.	n. s.
4	proportion of activity types	n. a.	n. a.	n. a.
6	total no. of writing incidents	n. s.	n. s.	n. s.
7	frequency of writing incidents	n. s.	n. s.	n. s.
8	mean duration of writing incidents	n. s.	n. s.	n. s.
9	mean length of writing incidents (char)	n. s.	n. s.	n. s.
10	total no. of revisions	n. s.	n. s.	n. s.
11	frequency of revising	n. s.	p<0.05	n. s.
12	proportions of revisions per TP phase	n. s.	n. s.	n. s.
13a	mean length of deletions in revision phase (char)	n. s.	n. s.	n. s.
13b	mean length of insertions in revision phase (char)	n. s.	n. s.	n. s.
14	total no. of information searches	n. s.	p<0.05	n. s.
15	frequency of information seeking	n. s.	n. s.	n. s.
16	mean duration of information searches	p<0.05	n. s.	n. s.
17	ratio of single-step to multi-step information searches	n. s.	n. s.	p<0.05
18	types of used online resources	n. a.	n. a.	n. a.
19	total no. of pauses	n. s.	n. s.	n. s.
20	frequency of pausing	n. s.	p<0.001	n. s.
21	mean duration of pauses	n. s.	n. s.	n. s.

List of all the measures and the results for intergroup comparisons, in sequential order as reported in this Chapter

Appendix D: Results for translator's products

This appendix contains all the translation products, that is the target texts produced by the 18 translators in the 24 translation processes. It also provides information on the preparation and analysis of data to investigate these translation products, which was done in the substudy (see Section 4.3 of the chapter *Methodology*). It also reproduces all the results that are presented in Chapter 6 *The translators' products*.

D1 Target texts BiDir 1-6 L1-L2 (English)

BiDir1 into L2

Stranding Risk of Whales
It is believed that the reason for this phenomenon is not due to an inclination to commit suicide. It seems to be rather a couple of factors that also may vary from case to case that cause and encourage the stranding of whales. The best researched strandings are those of the beaked whale. Those strandings seem to be related to the use of certain sonar types. After the use of these sonar types a stranding pattern has been observed that does not correlate with the normal stranding pattern of those whales: a lot of beaked whales stranded within a few hours, spread over several kilometres along the coast. With some whales the scientists found injuries of the acoustic organ which points towards a loss of the ability to navigate.

BiDir2 into L2

The stranding of whales
The phenomenon is unlikely to be linked to an inclination to suicide. Rather, there are probably several – and often very individual – factors causing or facilitating the stranding of live whales. The best-examined strandings so far are those of beaked whales, for which an association with the use of certain kinds of sonars is suspected. After the use of sonars, a stranding pattern unusual for this species has repeatedly been observed. Many beaked whales stranded within a few hours only, scattered over many kilometres of coastline. In many of them, the researchers found lesions in the acoustic organs, indicating a loss of the whales' ability to navigate.

(number of words: 112)

BiDir3 into L2

Whale strandings
The phenomenon is unlikely to be attributed to an inclination to commit suicide. It's rather a combination of several factors that cause or favour strandings of living whales, and these may differ from case to case. The strandings investigated in the greatest detail are those of beaked whales, which are suspected of being related to the use of certain types of sonar. An unusual stranding behaviour has often been observed for this species after sonar operations: a great number of beaked whales stranded within only a few hours, scattered across a coastline of several kilometres. Many of them were found to have injuries of their acoustic organs, which researchers attributed to a loss of the whales' navigation abilities.

BiDir4 into L2

Beachings of whales
Most likely, the phenomenon is not the result of an inclination to suicide. Rather, the factors causing or favoring the beachings of live whales are manifold and differ from one case to the other. The beachings investigated most thoroughly are those involving beaked whales, for which a connection to the use of certain types of sonar appears to exist. Following instances of sonar employed in this way, researchers frequently observed a beaching pattern untypical of this species: Many beaked whales stranded on land within a period of a few hours, spread over many miles of shoreline. In some of them, scientists detected injuries to the acoustic organs that suggested loss of the whales' ability to navigate.

BiDir5 into L2

Beached whales
It is not very likely that whales have a penchant to commit suicide. It is much more likely that there are often multiple factors that may cause or assist the beaching of living whales and that they differ from case to case. The beachings that have been researched most thoroughly are those of beaked whales, which are supposed to be connected to the use of specific types of sonar. After this kind of sonar had been used, a stranding pattern unusual for this genus was observed various times. A great number of beaked whales beached within a few hours, scattered along many kilometres of the coastline. Researchers detected that some of these whales had injuries to their organs of hearing which suggest that they lost their navigational abilities.

BiDir6 into L2

Whale beachings
An inclination for suicide is probably not the reason for this phenomenon. It is much more likely that many different factors varying case-by-case cause or facilitate beachings of living whales. The beachings best studied are those of beaked whales, where a connection is suggested with the use of certain sonar types. After such sonar operations, scientists were able to observe repeatedly a stranding pattern unusual for this genus: Many beaked whales stranded within few hours and were scattered over many miles of coastline. Scientists detected injuries to the hearing organs on some of them, which suggests a loss of the ability of navigation.

D2 Target texts BiDir 1-6 L2-L1 (German)

BiDir1 into L1

Sonarsystem gefährdet Wale
Kürzlich hat ein US Richter der amerikanischen Marine verboten, ein Sonarsystem zu testen. Dieses System ähnelt dem System welches das Verteidigungsministerium nur zu gerne einführen möchte. Der Richter befand, dass die dröhnenden Geräusche die Meeresbewohner beeinträchtigen und schädigen könnten. Sein Urteil hält Grossbritannien allerdings nicht davon ab, dem Niederfrequenz-Rennen beizutreten bei welchem leistungsfähige Lautsprecher auf Metallpfosten ins Meer versenkt werden. Eine sehr laute Geräuschabfolge, welche dazu dient feindliche Schiffe aufzuspüren, überflutet den Ozean. Dies verursacht Panik unter den Walen, welche bei der Nahrungs- und Partnersuche ähnliche, sonare Geräusche verwenden.

BiDir2 into L1

Wale durch Sonareinsatz bedroht
Ein amerikanischer Richter hat kürzlich der US-Navy untersagt, ein ähnliches System zu testen, wie jenes, welches das britische Verteidigungsministerium gern einführen möchte. Obwohl der Richter zu dem Schluss kam, dass das dröhnende Geräusch der Flora und Fauna des Meeres schaden könne, hält dies die Briten nicht davon ab, die Niederfrequenztechnologie anwenden zu wollen, bei der an einem Metallrohr angebrachte leistungsstarke Lautsprecher ins Meer hinabgelassen werden. Ein intensiver Lärmimpuls, mit dem feindliche Schiffe aufgespürt werden sollen, überflutet die Meere

und versetzt Wale in Panik, die für die Futter- und Partnersuche ähnliche Schallwellen nutzen.

(Anzahl Wörter: 96)

BiDir3 into L1

Gefahr für Wale durch Sonarübungen
Der amerikanischen Navy wurde kürzlich durch einen Richter untersagt, ein ähnliches System zu testen, welches auch das Verteidigungsministerium gerne einsetzen würde. Der Richter kam zum Schluss, dass die lauten Schallwellen für die Meereslebewesen schädlich sein könnten. Diese Bedenken konnten jedoch Grossbritannien nicht davon abhalten, ins "Niederfrequenzrennen" mit einzusteigen, wo mächtige Lautsprecher an einer Metallstange ins Meer herabgelassen werden. Solche Lautsprecher überfluten die Meere mit riesigen Mengen an Schallwellen, um gegnerische Schiffe zu orten. Dies führt zu Panik unter den Walen, welche auf der Suche nach Nahrung oder paarungsbereiten Partnern ähnliche Schallwellen aussenden.

BiDir4 into L1

Wale gefährdet durch martime Sonarübungen
Jüngst untersagte ein US-Richter der US-amerikanischen Marine das Testen eines Systems, das demselben gleicht, dessen Einführung das britische Verteidigungsministerium eifrig betreibt. Der Richter folgerte, dass die dröhnenden Geräusche die Meeresfauna und -flora schädigen könnten, doch seine Ausführungen haben wenig ausgerichtet, Grossbritannien vom Einstieg in das Niederfrequenzrennen abzuhalten, im Rahmen dessen leistungsstarke Lautsprecher an einem Metallstab in die See eingetaucht werden. Bei diesem Verfahren 'überflutet' ein zur Erfassung feindlicher Schiffe entwickelter intensiver Geräuschimpuls das Meer und verursacht Panik unter Walen, die sich ähnlicher Schallimpulse bedienen, um Nahrung und Paarungspartner zu finden.

(Anzahl Wörter: 95)

BiDir5 into L1

Wale in Gefahr wegen militärischen Sonar-Übungen im Meer
Kürzlich verbot ein US-amerikanischer Richter der US-Marine das Testen eines Systems, das demjenigen ähnelt, welches das britische Verteidigungsministerium unbedingt einführen möchte. Der Richter kam

zum Schluss, dass das Dröhnen den Meeresbewohnern schaden könnte, jedoch hat sein Kommentar in Grossbritannien keine grossen Wellen geschlagen – der Inselstaat will trotzdem ins Niederfrequenz-Rennen einsteigen, bei dem leistungsstarke Lautsprecher an einer Metallstange ins Meer hinabgelassen werden. Ein intensives, explosionsartiges Geräusch, das dazu dient, feindliche Schiffe zu entdecken, dröhnt durch den Ozean und löst unter den Walen Panik aus, da diese bei der Nahrungs- und Partnersuche ähnlich dröhnenden Schall verwenden.

BiDir into L1

Wale in Gefahr aufgrund von Sonar-Tests im Meer
Unlängst verbot ein US-amerikanischer Richter der US-Navy Tests mit einem ähnlichen System durchzuführen, das das britische Verteidigungsministerium gerne einsetzen möchte. Der Richter befand, dass die dröhnenden Laute für Meereslebewesen schädlich sein könnten, doch seine Einwände konnten nicht verhindern, dass Grossbritannien weiterhin ein Niederfrequenz-Sonarsystem einsetzen will, bei dem leistungsstarke Lautsprecher auf Metallpfosten ins Meer gelassen werden. Ein enormer Lärmimpuls zur Erkennung von Wasserfahrzeugen durchflutet den Ozean und löst bei den Walen damit Panik aus. Die Wale benutzen ähnliche Dröhngeräusche um Futter und mögliche Paarungspartner zu finden.

D3 Target texts UniEnglish 1-6 L2- L1

UniEnglish1 into L1

Beached whales
Suicidal tendencies are not thought to be behind this phenomenon. In fact, it is more often various factors that frequently differ from case to case which cause or favour the beaching of live whales. The beachings which have been the subject of most research are those of beaked whales, which have been linked to the use of specific types of sonar. After the use of such sonar, an unusual tendency to beach has been observed in this species. Many beaked whales beach within a few hours over many kilometers of coastline. Researchers have found injuries to their acoustic organs in some cases, which point to a loss of ability to navigate.

UniEnglish2 into L1

Stranded whales
The phenomenon is hardly likely to be due to a death wish on the part of the whales. Most often several factors are at work that cause whales to strand or increase the likelihood of their doing so, and these can vary from one incident to another. The most thoroughly investigated strandings are those of Cuvier's beaked whales where a link is thought to exist with the use of particular types of sonar technology. Numerous stranding patterns were observed following the use of such sonar technology that were untypical for this particular species. A large number of beaked whales stranded within a few hours, spread over a considerable length of coastline. In many of the stranded animals researchers observed injuries to their hearing organs, which indicated that they had lost their ability to navigate.

UniEnglish3 into L1

Beached whales
This phenomenon is probably not due to suicidal tendencies. Instead, whale strandings seem mostly to be caused or favoured by different factors, which often vary from case to case. The beachings that have been investigated the most thoroughly are those of beaked whales. It is suspected that there is a connection between the beaching of these whales and the uses of certain kinds of sonar. Following the use of this sonar, beaked whales stranded in a pattern that is unusual for their species: a large number of them beached within a few hours along many kilometres of coast. The researchers discovered that the acoustic organs of some of those whales were damaged, which suggests that they lost their navigational skills.

(Anzahl Wörter: 96)

UniEnglish4 into L1

Beached whales
The phenomenon should not be put down to the intention of committing suicide: the stranding of living whales is more often attributed to several, and often on a case-by-case basis, varying factors. The best-investigated strandings are those of beaked whales and it is thought that these are linked to the use of certain types of sonar. Based on this kind of sonar use, an unusual stranding pattern has been repeatedly identified for the species: many beaked whales were stranded within just a few hours, spread across many kilometers of coastline. In a number of cases, researchers identified

injuries to the hearing organs of the animals, an indication that the whales lost their ability to navigate.

(Number of words: 96)

UniEnglish5 into L1

Whale beachings
It is not likely that this phenomenon is the result of suicidal tendencies. There are usually numerous factors that cause live whales to beach, which often differ from case to case. Beachings of beaked whales have been investigated in the most detail, and in these cases it is suspected that there is a link to the use of certain types of sonar. An unusual beaching pattern was often observed following the use of these sonar devices. Many beaked whales beached within a few hours, scattered along several kilometres of coastline. Researchers found some of them to have suffered damage to their hearing organs which led them to lose their ability to navigate.

UniEnglish6 into L1

Beaching of whales
It is unlikely that the phenomenon has anything to do with a propensity to commit suicide. Instead, it is thought to be several factors, which often differ from case to case, that cause or encourage live whales to beach themselves. It is the beaching of beaked whales that has been investigated in most depth; here it is suspected that the beachings may be connected to the use of certain types of sonar. After using sonar of this kind, an unusual beaching pattern has been observed for this species: a large number of beaked whales beach themselves within a few hours, spread out across several kilometers of coastline. In some cases, researchers have established injuries to the hearing organs, which suggests that the whales may lose the ability to navigate correctly.

D4 Target texts UniGerman 1-6 L2- L1

UniGerman 1 into L1

Tests mit Sonarsystemen im Meer: Gefahr für Wale
Kürzlich verbot ein US-amerikanischer Richter der amerikanischen Navy das Testen eines ähnlichen Systems wie es das US-Verteidigungsministerium gerne einführen möchte. Der Richter kam zum Schluss, dass die Knallgeräusche das Meeresleben zerstören könnten. Seine

Äusserungen konnten jedoch Grossbritannien nicht davon abhalten, in das Wettrennen um Niederfrequenz-Sonarsysteme einzusteigen, bei dem auf einem Metallpfosten befestigte leistungsstarke Lautsprecher ins Meer hinuntergelassen werden. Der heftige Geräuschimpuls, mit dessen Hilfe feindliche Schiffe geortet werden, durchdringt den Ozean. Dies verursacht Panik unter den Walen, die bei der Nahrungs- und Partnersuche ähnliche Schallwellen einsetzen.

UniGerman 2 into L1

Sonarübungen gefährden Wale
Ein US-Richter hat der amerikanischen Marine jüngst verboten, ein ähnliches System zu testen, wie das, welches das Verteidigungsministerium einführen möchte. Der Richter kam zum Schluss, dass die dröhnenden Geräusche das Leben im Meer beeinträchtigen könnten. Allerdings haben seine Kommentare Grossbritannien nicht davon abgehalten, ebenfalls ins Niedrigfrequenzrennen einzusteigen. Dabei werden an einer Metallstange leistungsstarke Lautsprecher ins Meer hinuntergelassen. Diese dienen dazu, den Ozean mit intensivem Lärm zu beschallen, um feindliche Schiffe zu orten. Dieser Lärm verursacht Panik unter den Walen, da diese ähnliche Sonargeräusche verwenden, um Nahrung und paarungswillige Partner zu finden.

UniGerman 3 into L1

Marine Sonartests als Gefahr für Wale
Der US-amerikanischen Marine würde jüngst gerichtlich untersagt, ein System zu testen, wie es das britische Verteidigungsministerium einführen möchte. Der Richter begründete seinen Entscheid damit, dass die dröhnenden Geräusche des Systems die marine Fauna beeinträchtigen könnten. Dieses Urteil konnte allerdings nichts daran ändern, dass die Briten unbeeindruckt an ihrem Vorhaben festhalten, den Einsatz von niederfrequenten Sonargeräten voranzutreiben. Dabei werden leistungsstarke Lautsprecher an Metallstangen ins Meer hinabgelassen, mit dem Ziel, durch massive Lärmemissionen feindliche Schiffe zu erkennen. Dieser Lärm erzeugt jedoch Panik unter Walen, die auf Nahrungs- und Partnersuche ähnliche Laute einsetzen.

(Anzahl Wörter: 95)

UniGerman 4 into L1

Sonarübungen im Meer gefährden Wale
Vor Kurzem hat ein US-Richter der amerikanischen Marine verboten, ein ähnliches System zu testen, welches das MoD gerne einführen würde. Der Richter kam zum Schluss, dass die anschwellenden Töne das Leben im Meer gefährden könnten. Seine Kommentare vermochten Grossbritannien indes nicht davon abhalten, in das Niedrigfrequenzrennen einzusteigen, bei welchem effiziente Speaker auf einem Metallpfeiler in das Meer abgesenkt werden. Ein gewaltiger Lärm, der feindliche Schiffe aufspüren/orten sollte, breitet sich über den Ozean aus und führt zu Panikattacken unter den Walen, die bei der Nahrungs- und Partnersuche ähnliche Sonargeräusche ausstossen.

UniGerman 5 into L1

Sonarsystem der US-Marine gefährdet die Wale
Ein US-Richter hat jüngst ein Verbot ausgesprochen, das es der Navy unter anderem verbietet, ein vom Verteidigungsministerium geplantes Sonarsystem einzuführen. Der Richter begründet den Entscheid damit, dass die dumpfen Lärmemissionen die Lebenswesen im Meer gefährden könnten. Trotzdem arbeitet die britische Marine weiter an der Entwicklung von Niedrigfrequenz-Sonarsystemen. Dabei werden leistungsstarke Lautsprecher auf grossen Metallpfeilern befestigt und im Meer versenkt. Der laute Unterwasserknall dient dazu, feindliche Schiffe an der Wasseroberfläche zu orten. Die Wale reagieren mit Panik auf solche Störgeräusche, senden sie doch bei der Suche nach Nahrung oder Paarungspartnern ähnliche Geräusche aus.

UniGerman 6 into L1

Risiken für Wale bei Sonarübungen im Meer
Kürzlich hat ein US-Richter der Navy verboten, ein System zu testen, welches das Verteidigungsministerium in ähnlicher Form auch gerne einführen würde. Der Richter kam zum Schluss, dass die lärmenden Geräusche die Meeresfauna und -flora beeinträchtigen könnten. Seine Schlussfolgerungen haben jedoch wenig dazu beigetragen, Grossbritannien davon abzuhalten, ebenfalls Geräte im Niederfrequenzbereich einzusetzen; leistungsstarke Lautsprecher werden hierbei an einer Metallstange ins Meer herab gelassen. Dabei werden laute Geräusche ins Meer entlassen, um feindliche Schiffe aufspüren zu können; dies führt zu Panik unter den Walen, welche ähnliche Schall-Geräusche machen, um Nahrung und Paarungspartner zu finden.

D5 Data preparation and data analysis of substudy

Flesch Reading Ease scores of English target texts

I ensured that the online tool[163] I wanted to use produced correct results for the reading ease scores by doing some step-by-step calculations with the original formula. In general, results depend on how colons and semi-colons are considered – whether as full sentence delimiters or not. As in the English TTs, 7 of 12 contained at least one of these punctuation marks, treating them differently might have had an impact on results. As recommended by Flesch (n. d.), I treated colons and semi-colons as delimiters if they were followed by a complete sentence with a predicate. Titles were excluded from the calculation.

Statistical measure to compare product quality

In order to be able to choose the appropriate statistical test to check whether there existed a difference in quality between the English TTs produced by Bidir_L2 and those produced by UniEnglish_L1, some characteristics of the data had to be determined. Based on descriptive statistics and their visualisation as boxplots, it turned out that the data was not normally distributed and equal variances could not be assumed. Therefore, I decided to use the non-parametric two-sided Welch's *t*-test conducted on the ranks of the data. The results for product quality are presented in Section 6.2.

Raters' evaluation criteria

The 36 raters listed 108 criteria they had applied in evaluating the twelve English target texts. All of them had listed at least one criterion, 24 mentioned three criteria or more. The order in which they were listed is not taken into account. I grouped them into five categories without taking the order in which they were listed into account. The following table lists the categories and illustrates with two examples of how the participants phrased their criteria. For the questionnaire given to the raters see Appendix B.

163 http://www.webpagefx.com/tools/read-able/.

evaluation category	examples of original wordings
readability and comprehensibility	easy to read, easy to understand
grammar and syntax	punctuation, sentence structure
idiomatic expressions and word choice	collocations, word choice in 1st sentence
text composition and coherence	information structure, lack of continuity
text type adequacy	facts, style (scientific)

Categories of evaluation criteria and examples from raters

Although applied linguistics usually differentiates between the concepts of readability and comprehensibility (see e. g. Suojanen et al. 2015, 51–56), I decided to subsume them in a single category. I assumed that laypeople do not necessarily make that distinction and since 31 of the raters stated that they had had no experience in text revision or editing, they were considered laypeople in that respect. Out of the 36 raters only four (all of which were L2 raters and students of translation), reported evaluation criteria related to the fact that the texts were translations, such as for example 'transfer from German'.

Acceptability

The 36 raters evaluated the 12 target texts. They then ranked the three top ones in positions 1 to 3 and the three bottom ones in positions 10 to 12. Entries in positions 1 to 3 were rewarded while entries in positions 10 to 12 were penalised. The total Acceptability score was then calculated. For operationalisations see Section 4.3.6 and a description of results see Section 6.2.2

Target text	No. of mentions in 1st position	No. of mentions in 2nd position	No. of mentions in 3rd position	Score for positions 1-3	No. of mentions in 10th position	No. of mentions in 11th position	No. of mentions in 12th position	Score for positions 10-12	Total Acceptability score
Bidir1_L2	0	2	3	52	2	7	4	145	-93
Bidir2_L2	1	1	2	43	2	5	2	99	-56
Bidir3_L2	2	4	1	78	3	2	2	76	2
Bidir4_L2	3	1	2	67	4	4	5	144	-77
Bidir5_L2	4	4	2	112	3	2	3	88	24
Bidir6_L2	1	2	7	104	5	2	4	120	-16
UniEnglish1_L1	2	5	0	79	2	3	4	101	-22
UniEnglish2_L1	1	5	3	97	9	3	6	195	-98
UniEnglish3_L1	8	0	6	156	2	1	0	31	125
UniEnglish4_L1	5	4	2	124	1	2	3	68	56
UniEnglish5_L1	5	2	7	152	0	2	2	46	106
UniEnglish6_L1	4	6	1	124	3	3	1	75	49

Results of ranking tasks by total rater group and calculation of acceptability score.

position	score
1	12
2	11
3	10
4–9	0
10	-10
11	-11
12	-12

Scores assigned to the positions in the ranking task to determine product acceptability

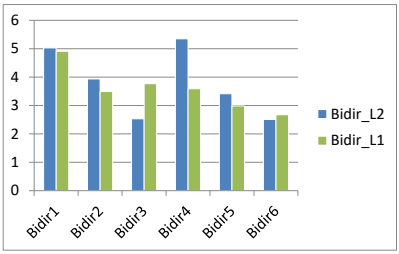
D6 Global measures of translation products

Number of target text characters and words

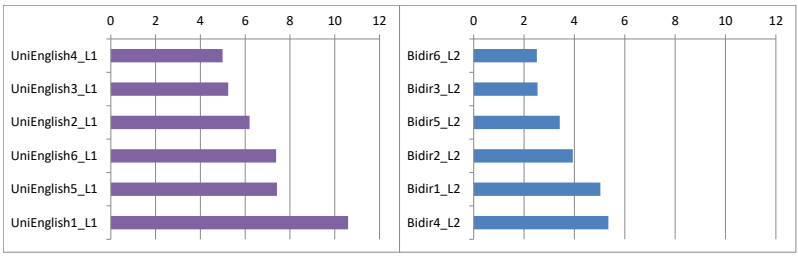
Target text	TT length (char)	TT length (words)	Target text	TT length (char)	TT length (words)
Bidir1_L2	757	133	UniEnglish1_L1	652	113
Bidir2_L2	686	112	UniEnglish2_L1	821	136
Bidir3_L2	743	120	UniEnglish3_L1	738	122
Bidir4_L2	730	119	UniEnglish4_L1	715	117
Bidir5_L2	763	130	UniEnglish5_L1	674	114
Bidir6_L2	663	105	UniEnglish6_L1	789	133
Bidir1_L1	732	93	UniGerman1_L1	757	94
Bidir2_L1	702	96	UniGerman2_L1	737	94
Bidir3_L1	743	97	UniGerman3_L1	758	96
Bidir4_L1	778	95	UniGerman4_L1	685	94
Bidir5_L1	790	102	UniGerman5_L1	759	98
Bidir6_L1	726	93	UniGerman6_L1	756	99

Global measures of translation products: number of characters and number of words in target texts

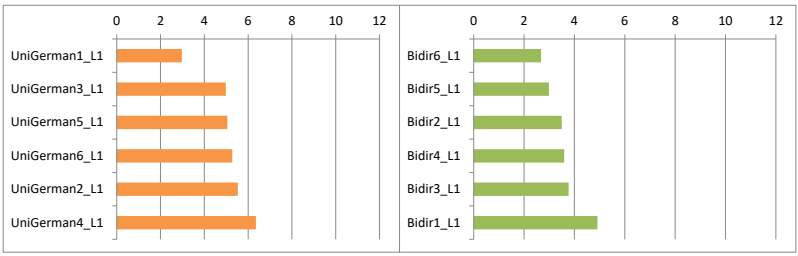
Productivity



Productivity for the bidirectional translators for the translation processes into L2 and into L1, ordered by translator (words/min)



Productivity for UniEnglish_L1 and Bidir_L2 groups, in ascending order (words/min)



Productivity for UniGerman_L1 and Bidir_L1 groups, in ascending order (words/min)

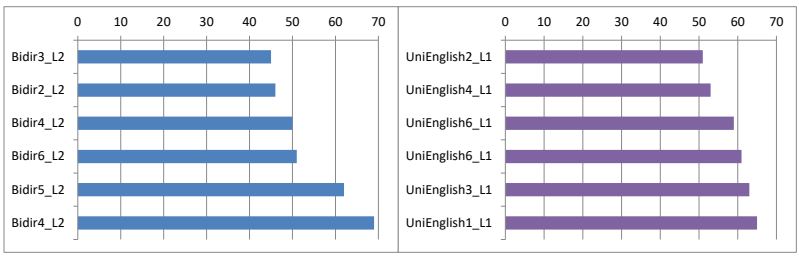
D7 Product quality of English target texts

Product quality was only determined for the English texts produced by the bidirectional translators into their L2 and by the English unidirectional translators into their L1. Readability and acceptability were used as indicators of product quality. For operationalisations see Section 4.3.6 and for description of results see Section 6.2.

Target text	Acceptability score by *total rater group* (n=36)	Acceptability score by subgroup *L1 raters* (n=18)	Acceptability score by subgroup *L2 raters* (n=18)
Bidir1_L2	-93	-36	-57
Bidir2_L2	-56	-56	0
Bidir3_L2	2	-24	26
Bidir4_L2	-77	-45	-32
Bidir5_L2	24	57	-33
Bidir6_L2	-16	23	-39
UniEnglish1_L1	-22	10	-32
UniEnglish2_L1	-98	-84	-14
UniEnglish3_L1	125	58	67
UniEnglish4_L1	56	10	46
UniEnglish5_L1	106	31	75
UniEnglish6_L1	49	56	-7

Acceptability scores for TTs of Bidir_L2 and UniEnglish_L1 by rater groups, ordered by translator

Readability of the English target texts



Flesch Reading Ease scores for English TTs of Bidir_L2 and UniEnglish_L1 groups, in ascending order

	Product quality (by *total rater group*)	Product quality (by *L1 raters*)	Product quality (by *L2 raters*)
Readability	0.121	0.351	-0.137

Rank correlations between readability and product quality of the English target texts (Kendall's τ)

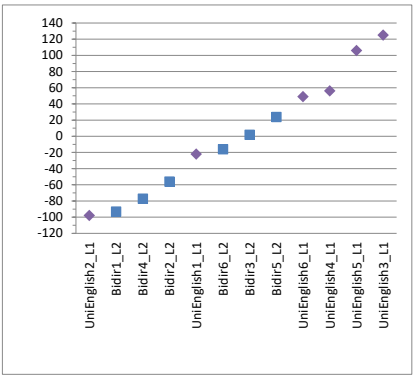
English newspapers	Flesch Kincaid Reading Ease score (Flaounas et al. 2012, 110)
The Guardian	30-35
The Wallstreet Journal	40-45
Independent	45-50
The Times	45-50
New York Times	45-50
New York Post	50-55
Daily Mirror	55-60
The Sun	60-65

Readability of English language newspapers as discussed in Section 6.2.1

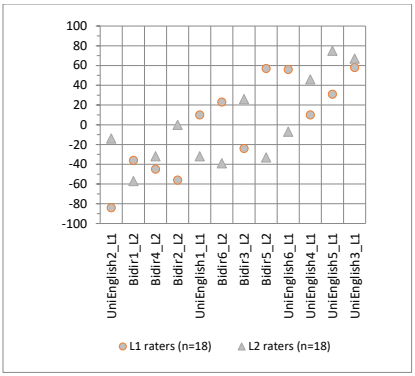
English target text	Flesch Kincaid Reading Ease score (calculated on http://www.webpagefx.com/tools/read-able/)
Bidir1_L2	69,4
Bidir2_L2	45,6
Bidir3_L2	44,9
Bidir4_L2	50,1
Bidir5_L2	62,2
Bidir6_L2	50,5
UniEnglish1	64,6
UniEnglish2	51,0
UniEnglish3	62,5
UniEnglish4	52,8
UniEnglish5	61,0
UniEnglish6	58,5

Readability of the English target texts

Acceptability of the English target texts



Acceptability scores for TTs of Bidir_L2 and UniEnglish_L1 in ascending order, based on evaluation by total rater group (n=36)



Acceptability scores for TTs of Bidir_L2 and UniEnglish_L1 based on evaluation by L1 raters and L2 raters, (in ascending order, based on evaluation by total rater group)

Raters' evaluation criteria

evaluation category	***total rater group* (n=36)**	**subgroup *L1 raters* (n=18)**	**subgroup *L2 raters* (n=18)**
readability and comprehensibility	24	29	20
grammar and syntax	24	23	25
idiomatic expressions and word choice	22	17	27
text composition and coherence	16	13	18
text type adequacy	14	17	11

Self-reported evaluation criteria per rater group (%), in descending order by total rater group

Appendix E: Results for translators' effort

This appendix contains information on the preparation and analysis of data to investigate the effort for translation (see Section 4.4 of the chapter *Methodology*). It also contains all the results that are presented in Chapter 7 *The translators' effort*. For the reader's convenience and for the sake of completeness, comparisons with significant and non-significant results are reproduced here.

Indicator	**Definition**	**Process measure(s) used**	**Product measure used**
Character production effort	Ratio of number of characters in process to number of characters in final product	Character count	Character count
Revision effort	Number of revisions per 10 TT words	Total number of revisions	Word count
Information integration effort	Number of search queries per 10 TT words	Total number of information searches	Word count
Coordination effort	Percentage of pauses with gaze focus on ST and TT	Total number of pauses. Representation of gaze data in SCR	n.a.
Mean fixation duration	Mean fixation duration during whole TP	Mean fixation duration	n.a.

Indicators of translator effort and their components

Comparisons of translators' effort between task groups

Indicator	**Bidir_L2 vs. Bidir_L1**	**Bidir_L2 vs. UniEnglish_L1**	**Bidir_L1 vs. UniGerman_L1**
Character production effort	$p<0.05$	n.s.	n.s.
Revision effort	n.s.	n.s.	n.s.
Information search effort	n.s.	$p<0.05$	n.s.
Coordination effort	n.s.	n.s.	n.s.
Mean fixation duration	n.s.	n.s.	n.s.

Overview of comparisons of translator effort indicators (significance level indicated; otherwise non-significant)

E1 Data preparation and data analysis of translator effort

Coordination effort

To investigate this, I subdivided the focus of attention into the following five categories (for operationalisation see Section 4.4 in *Methodology* chapter): focus on both the source and the target text, focus on the source text only, focus on the target text only, focus on the translation brief, and focus unclear. As data source, I used the screen recordings that contained the visualisation of eye-gaze behaviour. I found that the category *focus on ST and TT* tends to have the largest proportion of all foci during pauses. To measure coordination effort, I then calculated the ratio of the category focus on ST and TT.

Process	No. of pauses > 5 sec	Gaze focus during pauses (%)				
		On ST and TT	On ST only	On TT only	On translation brief	Unclear
Bidir1_L2	26	62	27	4	0	8
Bidir2_L2	30	40	43	13	0	3
Bidir3_L2	48	33	15	38	4	10
Bidir4_L2	20	65	10	15	0	10
Bidir5_L2	32	47	25	22	0	6
Bidir6_L2	32	44	13	34	3	6
Bidir1_L1	26	62	23	4	0	12
Bidir2_L1	31	55	32	13	0	0
Bidir3_L1	36	50	14	19	3	14
Bidir4_L1	20	45	15	15	5	20
Bidir5_L1	40	65	23	5	8	0
Bidir6_L1	25	44	24	24	4	4
UniEnglish1_L1	16	31	63	0	0	6
UniEnglish2_L1	44	50	2	36	5	7
UniEnglish3_L1	35	51	23	17	0	9
UniEnglish4_L1	28	29	43	11	4	14
UniEnglish5_L1	18	67	17	6	0	11
UniEnglish6_L1	22	50	18	23	5	5
UniGerman1_L1	37	54	30	5	3	8
UniGerman2_L1	23	74	9	9	0	9
UniGerman3_L1	37	41	32	24	0	3
UniGerman4_L1	20	50	40	5	0	5
UniGerman5_L1	24	25	33	42	0	0
UniGerman6_L1	40	48	35	8	3	8

Categories of gaze focus during pauses for each translation

Quality of eye-tracking data

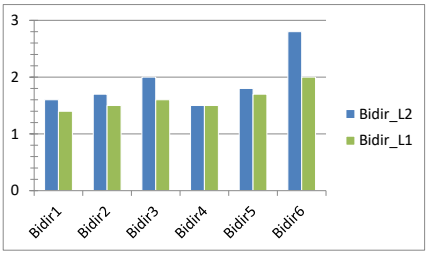
In my study, two of the translators have a mean fixation duration of less than 200 ms (UniGerman4 and UniGerman5) that is considered the threshold for a normal mean fixation length in translation processes recommended by Hvelplund (2011). Sharmin et al. (2008)[164] found that student translators who were touch typists had significantly lower average fixation durations than non-touch typists during translation tasks. They argue that since touch typists have more on-screen gaze time, which allows them to be more focussed, their fixations can be shorter. However, of the UniGerman_L1 group, only UniGerman3, whose mean fixation duration was within the common range, is not a touch typist. However, the two translators with mean fixation durations below 200 ms as well as the one with the highest duration are self-reported touch typists. As the calculation for mean fixation duration for UniGerman4 and UniGerman5 is based on 1647 and 1328 fixations during 85 % and 81 % of the translation process duration, respectively, and there is no evidence of erratic eye movements (such as reported as Brownian motion in Pavlović and Jensen 2009, 99), I decided to include the data in my sample.

164 Sharmin, S., Špakov, O., Räihä, K.-J., & Jakobsen, A. L. (2008). Where on the screen do translation students look while translating, and for how long? In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing (Vols 1–Book, Section, pp. 31–51). Frederiksberg: Samfundslitteratur.

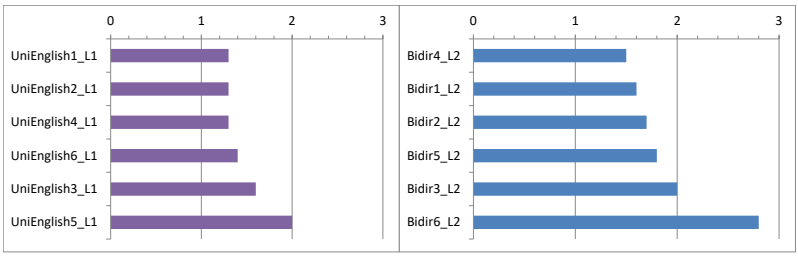
Process	No. of fixations	Mean fixation duration (ms)	Quality of eye-tracking data (%)
Bidir1_L2	2732	260	76
Bidir2_L2	3394	328	76
Bidir3_L2	5163	404	80
Bidir4_L2	1976	393	64
Bidir5_L2	4021	268	78
Bidir6_L2	4109	327	83
Bidir1_L1	1836	284	75
Bidir2_L1	3285	284	77
Bidir3_L1	2649	459	82
Bidir4_L1	2222	406	64
Bidir5_L1	3748	421	84
Bidir6_L1	3704	282	85
UniEnglish1_L1	1343	395	87
UniEnglish2_L1	2427	278	73
UniEnglish3_L1	2168	379	70
UniEnglish4_L1	2944	309	77
UniEnglish5_L1	1669	360	74
UniEnglish6_L1	1997	302	80
UniGerman1_L1	3616	456	90
UniGerman2_L1	695	208	51
UniGerman3_L1	2201	343	71
UniGerman4_L1	1647	194	85
UniGerman5_L1	1328	168	81
UniGerman6_L1	1553	661	91

Eye-tracking measures for all processes

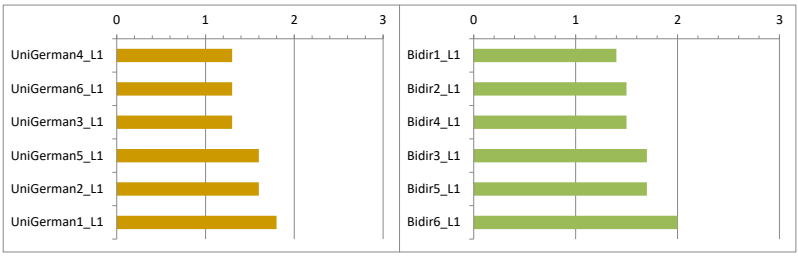
E2 Character production effort



Character production effort for Bidir_L2 and Bidir_L1, ordered by translator (process char/TT char)

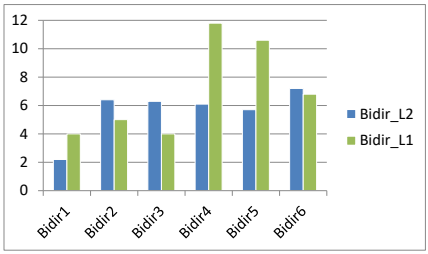


Character production effort for UniEnglish_L1 and Bidir_L2, in ascending order (process char/TT char)

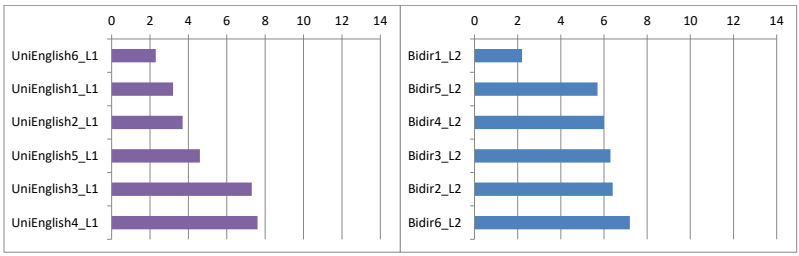


Character production effort for UniGerman_L1 and Bidir_L1, in ascending order (process char/TT char)

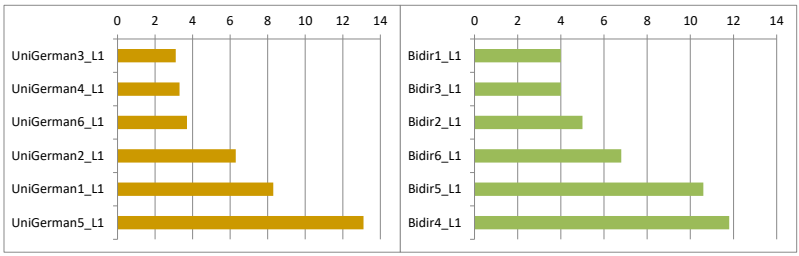
E3 Revision effort



Revision effort for Bidir_L2 and Bidir_L1, ordered by translator (revisions/10 TT words)



Revision effort for UniEnglish_L1 and Bidir_L2 in ascending order (revisions/10 TT words)



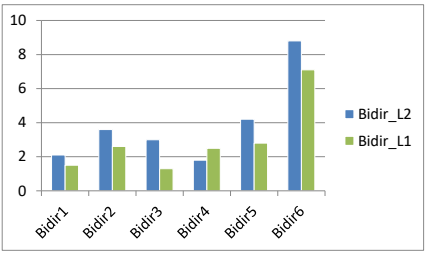
Revision effort for UniGerman_L1 and Bidir_L1 in ascending order (revisions/10 TT words)

E4 Information search effort

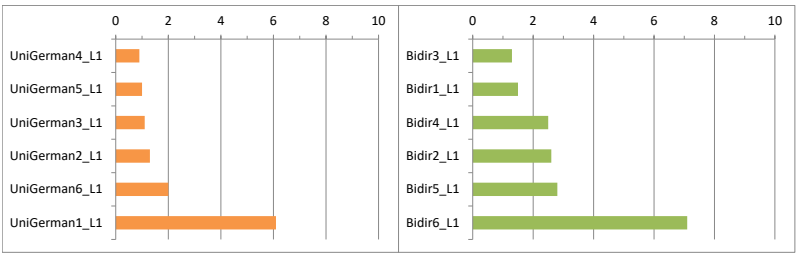
Process	Number of queries	Median	Process	Number of queries	Median
Bidir1_L2	28		Bidir1_L1	14	
Bidir2_L2	40		Bidir2_L1	25	
Bidir3_L2	36		Bidir3_L1	13	
Bidir4_L2	22		Bidir4_L1	24	
Bidir5_L2	55		Bidir5_L1	29	
Bidir6_L2	92	38	Bidir6_L1	66	24.5
UniEnglish1_L1	8		UniGerman1_L1	57	
UniEnglish2_L1	10		UniGerman2_L1	12	
UniEnglish3_L1	19		UniGerman3_L1	11	
UniEnglish4_L1	27		UniGerman4_L1	8	
UniEnglish5_L1	29		UniGerman5_L1	10	
UniEnglish6_L1	24	21.5	UniGerman6_L1	20	11.5

Number of search queries for all translation processes

Number of search queries

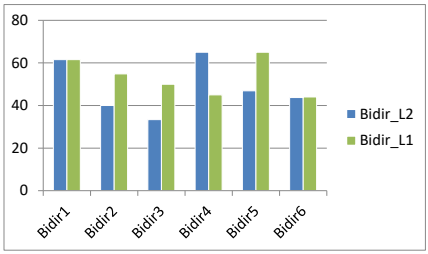


Information search effort for Bidir_L2 and Bidir_L1, ordered by translator (queries/10 TT words)

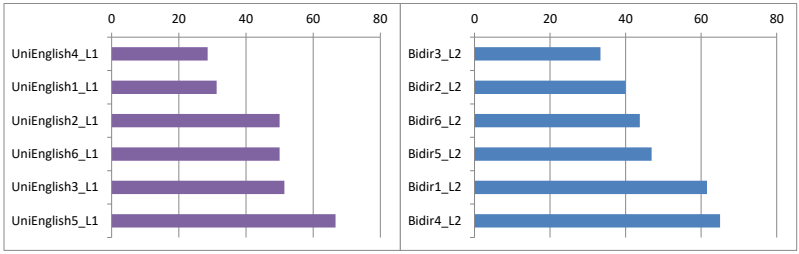


Information search effort for UniGerman_L1 and Bidir_L1, in ascending order (queries/10 TT words)

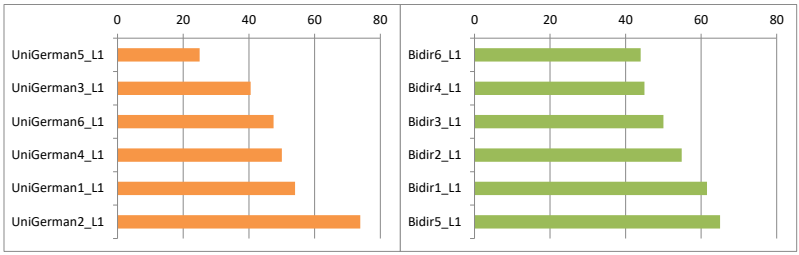
E5 Coordination effort



Coordination effort for Bidir_L2 and Bidir_L1 processes (%), ordered by translator

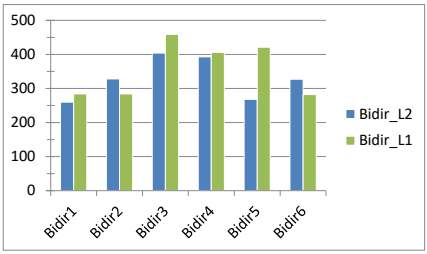


Coordination effort for UniEnglish_L1 and Bidir_L2 processes (%), in ascending order

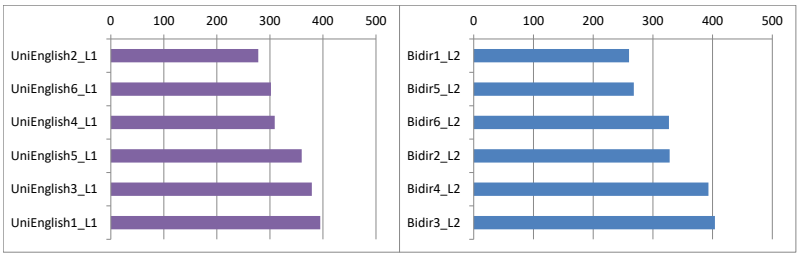


Coordination effort for UniGerman_L1 and Bidir_L1 processes (%), in ascending order

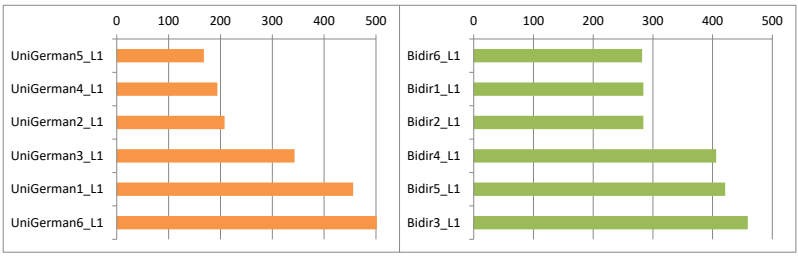
E6 Mean fixation duration



Mean fixation duration (ms) for Bidir_L2 and Bidir_L1 processes, ordered by translator



Mean fixation duration (ms) for UniEnglish_L1 and Bidir_L2 processes, in ascending order



Mean fixation duration (ms) for UniGerman_L1 and Bidir_L1 processes, in ascending order

Appendix F: Results for relations between characteristics of professional translation

This appendix reproduces the correlations performed to compare characteristics of professional translation as presented in Chapter 8 and introduced in Section 4.5. The characteristics consisted of translator effort, productivity and professional experience for all task groups. For the German-English task performed by the bidirectional translators into their L2 and the English unidirectional translators into their L1, product quality was added as a characteristic. In the respective correlations, product quality is referring to acceptability as operationalised in Section 4.3.6 and discussed in Section 6.2.2.

Correlation matrices with coefficients

Measure		Productivity		Professional experience	
		Bidir_L2	Bidir_L1	Bidir_L2	Bidir_L1
Translation effort	Character production effort	- - - - -	- - - -	- - - -	- - - -
	Revision effort	- - -	- - -	- - - -	- -
	Information integration effort	- - - -	- - - -	- - - - -	- - -
	Coordination effort	+ + + +	+ +	+ +	+
	Mean fixation duration	-	+ +	+	+
Productivity		n.a.	n.a.	+ + + +	+ + + + +

Results of correlations between translator effort, productivity and professional experience of the bidirectional translators (Kendall's τ)

Measure		Productivity		Professional experience	
		Bidir_L1	Uni-German_L1	Bidir_L1	Uni-German_L1
Translation effort	Character production effort	- - - -	- - -	- - - -	- -
	Revision effort	- - -	- -	- -	- -
	Information integration effort	- - - -	- - -	- - -	- -
	Coordination effort	+ +	+ +	+	+ +
	Mean fixation duration	+ +	- - -	+	- -
Productivity		n.a.	n.a.	+ + + + +	+ + + +

Results of correlations between translator effort, productivity and professional experience of the bidirectional translators into their L1 and the German unidirectional translators (Kendall's τ)

	Measure	Productivity		Professional experience		Product quality	
		Bidir_ L2	Uni-English_L1	Bidir_ L2	Uni-English_L1	Bidir_ L2	Uni-English_L1
Translation effort	Character production effort	- - - - -	-	- - - -	-	+ + +	+ + + +
	Revision effort	- - -	- - - -	- - - -	- -	+ +	+ +
	Information integration effort	- - - -	- -	- - - - -	- - -	+ + +	+ + +
	Coordination effort	+ + + +	+	+ +	+ +	- -	+ +
	Mean fixation duration	-	+ +	+	- -	+	+ + +
Productivity		n.a.	n.a.	+ + + +	-	- - -	- -
Professional experience		+ + + +	-	n.a.	n.a.	- - - -	- -

Results of correlations between translator effort, productivity, professional experience and product quality of the bidirectional translators into their L2 and the English unidirectional translators (Kendall's τ)

In der Reihe „Fach – Sprache – Kommunikation“ (ISSN 2699-7118) erschienen bisher folgende Titel:

Band 1
Franziska Heidrich & Klaus Schubert (Hrsg.): Fachkommunikation – gelenkt, geregelt, optimiert
Hildesheim: Universitätsverlag; 2020. – 190 S.
Freie Online-Publikation:
http://dx.doi.org/10.18442/087
Online-ISSN 2699-7118

One way to categorise translation is on the basis of translation direction, i. e. whether it is performed into or out of a translator's so-called native language. Both translation directions are widely practised around the globe. This empirical study in Cognitive Translation Studies investigated potential effects of translation direction on the translation process, the effort and the product. It focused on the German-English language pair and included qualified specialised translation professionals as participants. In a simulated workplace setting, their translation processes were monitored and analysed following a mixed methods approach. The translations were then compared applying an extensive set of process descriptors, novel indicators of translation effort and an innovative user-centred product evaluation. Results showed no statistically significant directionality effects for the vast majority of the process descriptors and effort indicators or for product quality.